UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE PFIZER INC. SECURITIES LITIGATION	No. 04-cv-9866 (LTS) (DFE)

[PROPOSED] FINDINGS OF FACT AND CONCLUSIONS OF LAW WITH RESPECT TO:

(1) PFIZER DEFENDANTS' MOTION TO EXCLUDE CERTAIN PLAINTIFFS' EXPERTS' OPINIONS REGARDING CELEBREX AND BEXTRA; AND

(2) PLAINTIFFS' MOTION TO EXCLUDE EXPERT TESTIMONY BY DEFENDANTS' EXPERT DR. LEE-JEN WEI

CADWALADER, WICKERSHAM & TAFT LLP
One World Financial Center
New York, New York 10281
Telephone: 212-504-6000
Facsimile: 212-504-6666

Of Counsel:

DLA PIPER LLP (US)
1251 Avenue of the Americas
New York, New York 10020-1104
Telephone: 212-335-4846
Facsimile: 212-335-4501

TUCKER ELLIS & WEST LLP
Metropoint 1, Suite 1325
4600 S. Ulster Street
Denver, Colorado 80237
Telephone: 720-897-4404
Facsimile: 720-222-5242

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Page

TABLE OF AUTHORITIES

Page

CASES

TABLE OF AUTHORITIES
(continued)

Page

TABLE OF AUTHORITIES
(continued)

Page

OTHER AUTHORITIES

Lead Plaintiff Teachers' Retirement System of Louisiana brought this action on behalf of a putative class of investors ("Plaintiffs") who purchased or acquired Pfizer Inc ("Pfizer") stock between October 31, 2000 and October 19, 2005 against Pfizer and certain of its corporate officers (collectively, "Defendants"). Plaintiffs allege that Defendants violated federal securities laws by concealing evidence of certain side effects related to two Pfizer medications, Celebrex and Bextra, and by making misstatements and omissions in their public filings and statements.

Now before the Court are: (1) Defendants' motion pursuant to Federal Rule of Evidence 702 to exclude the opinion of Plaintiffs' experts that prior to December 16, 2004, there existed reliable scientific evidence that Celebrex and/or Bextra was associated with a statistically significant increase in the risk of thrombotic events such as heart attacks and strokes; and (2) Plaintiffs' Rule 702 motion to exclude a meta-analysis conducted by one of the Defendants' experts, Dr. Lee-Jen Wei. The Court received extensive briefing on the motions and held a five-day hearing with testimony by certain of the experts, which was followed by additional written submissions. For the reasons outlined below, the Court hereby GRANTS the Defendants' motion and DENIES Plaintiffs' motion. This Opinion and Order constitutes the Court's findings of fact and conclusions of law in accordance with Federal Rule of Civil Procedure 52.

BACKGROUND

I. THE CELEBREX THROMBOTIC RISK WARNING IS BASED ON RESULTS FROM A SINGLE, EXPERIMENTAL, HIGH DOSE, CANCER PREVENTION TRIAL THAT HAS NOT BEEN REPLICATED IN ANY OTHER TRIAL.

Celebrex is the world's most widely prescribed non-steroidal anti-inflammatory drug ("NSAID") and is approved to treat the symptoms of osteoarthritis, rheumatoid arthritis, and other conditions. *See* Defs.' Ex. 630 at 3 (Celebrex Label, Dec. 2008). NSAIDs inhibit an enzyme called cyclo-oxygenase, or COX, which has two forms (COX-1 and COX-2). NSAIDs that inhibit both COX-1 and COX-2 are referred to as "non-selective" NSAIDs and include

ibuprofen (Advil and Motrin) and naproxen (Aleve); NSAIDs that primarily target COX-2 are referred to as "selective" COX-2 inhibitors and include Celebrex, Vioxx, and Bextra.

In 2000, a Vioxx trial known as VIGOR showed that patients taking Vioxx experienced a statistically significant, four-fold increased risk of heart attacks compared to patients taking prescription Aleve. *See* Defs.' Ex. 636 at 1520 (Bombardier et al., N. ENGL. J. MED., 2000;343:1520-28). The U.S. Food & Drug Administration ("FDA") analyzed the VIGOR data and found statistically significant evidence of an increased thrombotic risk for patients taking Vioxx in that trial, and subsequently supplemented the FDA-approved Vioxx label. *See* Defs.' Ex. 640 at 12 (Vioxx Label, Apr. 2002). Around the same time, a prominent medical doctor named Garrett FitzGerald was studying how NSAIDs work in the body and hypothesized that selective COX-2 inhibitors increase the risk of clotting events in the arteries, most notably heart attacks and strokes.[1] *See* Defs.' Ex. 1212 at 28D (FitzGerald, AM. J. CARDIOL. 2002;89:26D-32D). In contrast to VIGOR and Dr. FitzGerald's thrombotic risk hypothesis, the results of a very large Celebrex trial known as CLASS showed no evidence that patients taking four times the approved osteoarthritis dose had a statistically significant increased risk of thrombotic events compared to patients taking standard doses of non-selective NSAIDs, which the FDA noted in the Celebrex label. *See* Defs.' Ex. 639 at 21 (Celebrex Label, June 2002).

On September 30, 2004, Merck withdrew Vioxx from the market after another study, APPROVe, showed that patients taking Vioxx experienced a statistically significant, two-fold increase in thrombotic events compared to patients taking placebo. *See* Defs.' Ex. 638 at 1092

[1] Heart attacks and strokes are most often caused by clots that restrict the flow of blood to the heart or brain. Those events are referred to as "thrombotic" events to distinguish them from non-thrombotic cardiovascular problems that are not induced by clots, such as abnormal heart rhythms (arrhythmias) or the heart's inability to pump enough blood (congestive heart failure).

(Bresalier et al., N. ENGL. J. MED. 2005;352;1092-1102); Defs.' Ex. 1998 (Merck Press Release, Sept. 30, 2004).[2] Prior to the Vioxx withdrawal, Celebrex had been studied in trials involving more than 25,000 arthritis patients – more than 10,000 of whom took Celebrex every day for at least six months – none of which revealed a statistically significant increase in the risk of thrombotic events. *See* Defs.' Ex. 637 at 16-18 (Pfizer Briefing Doc.); Defs.' Br., App., Figs. 4 & 5. A series of observational studies comparing patients who took Vioxx or Celebrex to patients who took non-selective NSAIDs or no NSAID at all also consistently showed no increased risk of thrombotic events with Celebrex, but did show an increased risk with Vioxx.[3] At the same time, independent safety monitoring committees overseeing ongoing experimental, high-dose studies of Celebrex for the prevention of cancer and Alzheimer's disease reviewed the data from those clinical trials in light of the Vioxx withdrawal and found no reliable evidence that Celebrex increased the risk of heart attacks and strokes. *See* Defs.' Ex. 660 at 1 (Bertagnolli Mem., Oct. 15, 2004); Defs.' Ex. 661 at 1 (NIH Press Release, Dec. 20, 2004).

On December 16, 2004, the National Cancer Institute suddenly halted the Adenoma Prevention with Celecoxib trial ("APC"), which was designed to test whether Celebrex could prevent pre-cancerous colorectal tumors. That trial, which studied patients taking two (400 mg

[2] Experts in cardiology formulated the primary thrombotic endpoint in the trial, which included heart attacks, strokes, unstable angina, and other thrombotic events. *See* Defs.' Ex. 638 at 354. Cardiology experts diagnosed each thrombotic event included in the statistical analysis based on all available medical information for each patient and in accordance with written diagnostic criteria for each type of event, and all disagreements between cardiology experts were resolved in accordance with a pre-specified process. *See id.* at 354-55.

[3] Dr. David Graham, an FDA Medical Officer, conducted one of these observational studies and wrote in an internal FDA memorandum: "[W]e estimate that the increased [Vioxx] risk observed in this study would yield an excess of 27,785 cases of [acute myocardial infarction] and [sudden cardiac death] in the US over the years 1999-2003 . . . . These would have been avoided had [Celebrex] been used instead of [Vioxx]." Defs.' Ex. 641 at 10 (Graham Mem. to Seligman, Sept. 30, 2004); *see* Defs.' Br. at 22 & n.18 (citing Defs.' Exs. 643 through 647).

daily) and four (800 mg daily) times the recommended daily osteoarthritis dose of Celebrex (200 mg daily) every day for almost three years, showed that patients taking high doses of Celebrex experienced a higher rate of heart attacks and strokes than patients taking placebo pills. *See* Defs.' Ex. 658 at 1071, 1075 (Solomon S. et al., N. ENGL. J. MED. 2005;352:1071-80). The differences were statistically significant, though the difference between patients taking two times (400 mg daily) the approved osteoarthritis dose and those taking placebo was only borderline significant. *See id.* at 1075 (reflecting barely significant confidence intervals).[4]

Immediately, the FDA announced that APC was the first time researchers saw a statistically significant increase in thrombotic events with Celebrex compared to patients taking placebo: "Previous large studies of Celebrex, including clinical trials and epidemiology studies, have not suggested the sort of [cardiovascular] risk found in the NCI polyp study." Defs.' Ex. 659 at 2 (FDA Statement, Dec. 17, 2004); *see* Defs.' Ex. 664 at 4-5 (FDA Decision Mem.). In fact, after reviewing all the Celebrex data, Plaintiffs' expert Dr. Joel Bennett agreed:

> Q. In your mind, Dr. Bennett, the evolving scientific data first established that 400 milligrams of Celebrex confers an increased thrombotic risk as of December 2004, when the APC data became available, correct?
>
> A. That's right.

Defs.' Ex. 676 at 182 (Bennett Hr'g Test., *In re Bextra*, Oct. 9, 2007).[5] Plaintiffs' other experts,

[4] As with APPROVe, the APC event counts were reliable because cardiology experts formulated the primary endpoint, diagnosed the events based on written diagnostic criteria and all available medical information, and pre-specified a process to resolve disagreements between the experts. *See id.* at 1072-73; Defs.' Ex. 793 at 67 (Bennett Dep.); Hr'g Tr. at 193 (Dr. Furberg).

[5] *See* Defs.' Ex. 612 at 385-86 (Bennett Dep., *In re Bextra* ["I think the concern was raised in 2004, the end of 2004."]); *id.* at 103-104 (agreeing that the first evidence of an increased risk at 400 mg daily was the APC data); *id.* at 198-200 (indicating that he never has known a real-world Celebrex patient to take as much medication as the patients in APC). Even after APC, Dr. Bennett wrote that the thrombotic risk of selective COX-2 inhibitors other than Vioxx "remains to be established." Defs.' Ex. 665 at 1835 (WILLIAMS HEMATOLOGY (7th ed., 2007)).

Drs. Curt Furberg and Richard Kronmal, reached the same conclusion. *See* Hr'g Tr. at 179 (Dr. Furberg); *id.* at 386 (Dr. Kronmal); Defs.' Ex. 621 at 345-46 (Kronmal Dep.). Similarly, in the Celebrex product liability litigation, courts found no reliable evidence of an increased risk at the dose most commonly used by arthritis patients and held that APC was the only trial to demonstrate a statistically significant increase in thrombotic risk at the higher doses.[6]

Moreover, the results of the experimental, high dose, cancer prevention APC trial have not been replicated in any other Celebrex clinical trial. Two other high-dose, long-term, disease prevention trials – one studying the preventive effects of Celebrex on colon cancer (PreSAP), which used the same adjudication procedures as APC, and the other studying the preventive effects on Alzheimer's disease (ADAPT) – showed no statistically significant differences in thrombotic events between patients taking high doses of Celebrex and those taking placebo. *See* Defs.' Ex. 662 at 885, 887 (Arber et al., N. ENGL. J. MED. 2006;355:885-95); Defs.' Ex. 663 at e33 (ADAPT Research Group, PLOS CLIN. TRIALS 2006;1(7):e33).[7] The FDA recognized that the PreSAP and ADAPT results were inconsistent with the APC results. *See* Defs.' Ex. 664 at 4-5, 9 (FDA Decision Mem.). Subsequent cancer prevention and other experimental disease prevention trials also consistently have shown no evidence of an increased thrombotic risk. *See* Defs.' Ex. 678 at 2107, Table 3 (Solomon S. et al., CIRCULATION 2008;117:2104-13).

On February 18, 2005, the FDA convened an Advisory Committee meeting to try to answer the public health question of whether and to what extent NSAIDs increase the risk of

[6] *See In re Bextra & Celebrex Mktg. Sales Pracs. & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1181-82 (N.D. Cal. 2007) ("*In re Bextra I*"); *In re Bextra & Celebrex*, 2008 N.Y. Misc. LEXIS 720 at *27-29, 32-34 (Sup. Ct. N.Y. County Jan. 7, 2008) ("*In re Bextra II*").

[7] Drs. Bennett and Furberg criticized ADAPT because it lacked the same data collection and diagnosis procedures as APC and PreSAP. *See* Defs.' Ex. 617 at 101-03 (Furberg Dep.); Defs.' Ex. 682 at 1639 (Antman et al., CIRCULATION, 2007;115:1634-42 ["2007 AHA statement"]).

thrombotic events such as heart attack and stroke. Based on the totality of available data, the Advisory Committee voted almost unanimously to keep Celebrex on the market. *See* Defs.' Ex. 666 at 11 (Ad. Comm. Minutes [31 to 1]). Based on the APC data, it also found no evidence of an increased risk at the most commonly used dose of 200 mg daily, a marginally positive risk at 400 mg daily, and an increased risk at 800 mg daily. *See id.*

On April 6, 2005, the FDA issued a memorandum setting forth its recommendations and analysis for all NSAIDs. The FDA defined the relevant public health question as whether and to what extent the various NSAIDs increase the risk of thrombotic events, based in part on the results from the VIGOR, APPROVe, and APC trials and the FitzGerald hypothesis. *See* Defs.' Ex. 664 at 1, 4, 8-9 (FDA Decision Mem.). To answer that question, the FDA chose an endpoint widely accepted by medical doctors in the cardiovascular community—the Anti-Platelet Trialists' Collaboration (or "APTC") composite endpoint, which includes non-fatal heart attack, non-fatal stroke, and vascular death. *See id.* at 4 (noting that the FDA chose APTC because it is "a widely accepted endpoint in assessing the benefits and risks of a drug for [cardiovascular] outcomes"). The FDA then applied the APTC endpoint consistently across the Celebrex clinical trials, including each trial on which Plaintiffs rely here, and found that APC was the only trial to show a statistically significant increase in thrombotic risk. *See id.* at 4-5. The FDA did not cite any meta-analyses of clinical trials in its lengthy decision memorandum. *See id.*, *passim*.

Based on precautionary public health principles, the FDA did find that APC furnished enough evidence to conclude for the first time that Celebrex increases thrombotic risk, "at least at some dose, with reasonably prolonged use." *Id.* at 10. At the same time, the FDA determined that Celebrex was no less safe than any non-selective NSAID, including prescription Motrin or Aleve. *See id.* at 8, 10. Because Celebrex was no less safe than other NSAIDs, the FDA

recommended that all NSAIDs, including Celebrex, carry the same "boxed" warning that NSAIDs "may" increase the risk of thrombotic events such as heart attack and stroke. *See id.* at 14; Defs.' Ex. 1598 at 1 (Celebrex Label, July 2005).

Thereafter, researchers continued to study the thrombotic safety of Celebrex. Based on the same APTC endpoint used by the FDA, a peer-reviewed and well-respected meta-analysis of clinical trials – on which Plaintiffs' experts rely – did not show a statistically significant increase in thrombotic risk, whether Celebrex was compared to placebo or other NSAIDs, even when the APC data were included. *See* Defs.' Ex. 784 at 1304 (Kearney et al., BRIT. MED. J., 2006;332:1302-08 ["Kearney"]); Defs.' Ex. 682 at 1635 (2007 AHA statement [citing Kearney]); Hr'g Tr. at 170, 278.[8] Similarly, observational studies of millions of real-world patients taking Celebrex at common doses prescribed by their doctors show no increased risk for patients taking Celebrex compared to those taking another NSAID or no NSAID at all. *See* Defs.' Ex. 683 at 1638-39, 1642 (McGettigan et al., JAMA 2006; 296:1633-44 ["McGettigan"]).

II. THE THROMBOTIC RISK WARNING ON THE BEXTRA LABEL WAS BASED ON THE TRIALS OF HIGH DOSE, INTRAVENOUS PARECOXIB IN PATIENTS IMMEDIATELY AFTER UNDERGOING OPEN HEART SURGERY.

On November 16, 2001, the FDA approved Bextra and concluded that it was safe and effective when used to treat the symptoms of arthritis at a dose of 10 mg once daily and primary dysmenorrhea (painful menstrual cramps) at a dose of 20 mg twice daily, as needed. *See* Defs.' Ex. 687 at 1 (FDA Approvable Ltr. for Bextra, Nov. 16, 2001). None of the trials of oral Bextra pills showed a statistically significant increase in thrombotic risk among patients taking Bextra – at any dose – compared to patients taking another NSAID or no NSAID at all. *See* Defs.' Ex.

[8] *See also* Defs.' Ex. 1815 at 91-92 (White et al., AM. J. CARDIOL. 2007;99:91-98 [finding no statistically significant increase in the risk of the APTC endpoint for trials completed before APC]); Defs.' Ex. 1813 at 411 (White et al., AM. J. CARDIOL. 2003;92:411-18 [same]).

686 at 28-29 (Bextra Integrated Summary of Safety, Dec. 2000). No published meta-analysis of the Bextra arthritis trials ever has shown a statistically significant increase in the risk of heart attacks, strokes, or heart attacks and strokes combined, whether Bextra was compared to placebo or other NSAIDs. *See* Defs.' Br. at 36 & n.43 (citing Defs.' Exs. 702-705); Defs.' Ex. 612 at 522-24 (Bennett Dep., *In re Bextra*); Defs.' Ex. 793 at 199-200 (Bennett Dep.). Observational studies, including one by Dr. Graham of the FDA, similarly concluded that real-world arthritis patients taking approved doses of Bextra are at no greater thrombotic risk than patients taking other NSAIDs or no NSAID at all. *See* Defs.' Ex. 700 at 1382-83 (Solomon D. et al., ARTHR. RHEUM. 2006; 54:1378-89); Defs.' Ex. 822 at 10 (Singh et al.).

Prior to the FDA's approval of Bextra, Pfizer also sought approval for parecoxib, an experimental, intravenous form of Bextra. Pfizer conducted two trials among patients who needed relief from the excruciating pain that occurs immediately after coronary artery bypass graft ("CABG") surgery, known as the CABG surgery trials, in which patients received high-dose, intravenous parecoxib immediately after undergoing CABG surgery, followed by oral Bextra pills at doses four and eight times the approved arthritis dose. While the CABG-1 trial had only two groups of patients – one that received intravenous parecoxib followed by oral Bextra pills and another that received intravenous placebo and placebo pills – the CABG-2 trial also had a third group of patients who received intravenous placebo followed by oral Bextra pills, but no parecoxib ("the Bextra pill only group"). *See* Defs.' Ex. 691 at 1481 (Ott et al., J. THORAC. CARDIOVASC. SURG. 2003;125:1481-92 ["Ott"]) (CABG-1); Defs.' Ex. 696 at 1082 (Nussmeier et al., N. ENGL. J. MED. 2005;352:1081-91 ["Nussmeier"]) (CABG-2). Thus, only the Bextra pill only group in CABG-2 permitted a comparison of oral Bextra pills to placebo pills. Using a peer-reviewed thrombotic endpoint and a reliable data collection process, the

difference between the Bextra pill only group and the placebo group was not statistically significant. *See* Defs.' Ex. 696 at 1082-83, 1087, Table 3 (Nussmeier); Defs.' Ex. 621 at 334-35 (Kronmal Dep.); Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249). The FDA never approved parecoxib, in part because of its unique side effects, including potentially life-threatening drops in blood pressure known as hypotension. *See* Defs.' Ex. 690 at 2 (FDA Ltr.).[9]

As with Celebrex, the FDA concluded that there was no evidence that real-world patients taking Bextra pills for approved uses were at any greater risk of heart attacks and strokes, whether compared to placebo or other NSAIDs like Motrin or Aleve. As Dr. Graham of the FDA noted: "With [Bextra] . . . the information we have at this time suggests that the risk is not increased at doses of 20 mg or less." Defs.' Ex. 632 at 83 (Ad. Comm. Tr., Feb. 17, 2005); *see* Defs.' Ex. 666 at 12 (Ad. Comm. Minutes ["[T]he Committee felt that the evidence was very limited and it is difficult to extrapolate to a real life setting."]); *id.* at 11 (voting to keep Bextra on the market). The FDA recommended that Pfizer withdraw Bextra because of an increased risk of very rare skin reactions compared to other NSAIDs, not a unique cardiovascular concern that distinguished Bextra from other NSAIDs. *See* Defs.' Ex. 664 at 17 (FDA Decision Mem. [noting that its precautionary public health assumption that Bextra carried the same thrombotic risk as other NSAIDs "would not be sufficient to warrant withdrawal of Bextra since we have no data showing that Bextra is worse than other NSAIDs" with regard to thrombotic risk]).

III. PLAINTIFFS' ALLEGATIONS AND EXPERTS.

In their complaint, Plaintiffs do not contend that Defendants failed to disclose the APC results in a timely fashion. *See generally* Defs.' Ex. 818 (Compl.). Rather, Plaintiffs claim that

[9] The CABG-2 investigators also conducted a trial of high-dose parecoxib followed by high-dose Bextra pills in patients undergoing *non-cardiac* surgery, which showed no increased thrombotic risk. *See* Defs.' Ex. 699 at 519, 523 (Nussmeier et al., ANESTHESIOLOGY 2006;104:518-28).

Defendants knew about but failed to disclose the thrombotic risk described in the boxed label before December 2004.[10] Plaintiffs' complaint and their experts – especially their blood clotting and thrombosis expert, Dr. Bennett – also rely on the FitzGerald hypothesis,[11] which Drs. FitzGerald and Bennett (the primary proponent of the hypothesis for Plaintiffs in this litigation) have described as thrombotic in nature and as applying to heart attacks, strokes, and other clots in the arteries – not an arbitrary collection of events that are confined to the vicinity of one body part or that involve no direct relationship to clots.[12]

To support their claim that there was reliable, statistically significant evidence of an increased thrombotic risk associated with Celebrex and Bextra prior to December 16, 2004, Plaintiffs designated Drs. Bennett, Furberg, Kronmal, and David Madigan in their affirmative case. Drs. Kronmal and Madigan are statisticians who do not have medical degrees, are not board-certified in cardiology, and do not claim expertise in how Celebrex and Bextra work in the human body. *See* Defs.' Ex. 708 at 10-11, 70 (Madigan Dep., *Grutka v. Pfizer*); Defs.' Ex. 621 at 49, 65, 123-27 (Kronmal Dep.). Dr. Furberg is not board-certified in cardiology, never has been licensed to practice medicine in the U.S., has not prescribed a medication in more than thirty-five years, admits that he is not competent to practice medicine today, and does not know the current diagnostic criteria for cardiovascular conditions such as heart attacks or pulmonary

[10] *See id.* at 3 ("It is now clear that Pfizer and its senior management knew that from its initial approval, Celebrex should have been the subject of the 'black box' FDA warning it now carries regarding the substantial risk of cardiovascular harm that can be caused by using the drug."); *see also id.* at 4, 5, 42, 54, 57-58, 60-73, 75-84 (relying on the boxed warning for Celebrex).

[11] *See* Defs.' Ex. 618 at 1, 5 (Bennett Rep.); Defs.' Ex. 610 at 17-18, 21, 53 (Furberg Rep.); Defs.' Ex. 670 at 7, 11 (Baruch Rep.); Defs.' Ex. 621 at 51 (Kronmal Dep.); Defs.' Ex. 706 at 1 (Madigan Rep.); Hr'g Tr. at 543-44, 547-49 (Dr. Madigan).

[12] *See* Defs.' Ex. 1212 at 28D (FitzGerald, AM. J. CARDIOL. 2002;89:26D-32D); Defs.' Ex. 682 at 1636-37 (2007 AHA Statement); Hr'g Tr. at 232 (admitting the hypothesis is thrombotic).

edema. *See* Hr'g Tr. at 180-81, 191-92; Defs.' Ex. 713 at 49 (Furberg Baycol Dep.); Defs.' Ex. 617 at 151, 153, 182-83, 350 (Furberg Dep.). Like Drs. Kronmal and Madigan, Dr. Furberg does not purport to be an expert in how selective COX-2 inhibitors might affect the body's clotting mechanisms. *See* Hr'g Tr. at 181-82; Defs.' Ex. 617 at 23-24 (Furberg Dep.); Defs.' Ex. 616 at 60 (Furberg Dep., *Haslam v. Pfizer*). Plaintiffs also designated Drs. Lawrence Baruch, Plaintiffs' only board-certified expert in cardiology, and Nicholas Jewell as rebuttal experts only.

At the hearing, Plaintiffs did not make the cardiologist Dr. Baruch or the thrombosis expert Dr. Bennett available for cross-examination, even though they are Plaintiffs' only board-certified physicians who treat clotting conditions and prescribe medications.[13] Plaintiffs' failure to present those experts speaks volumes about the methodologies used by the experts they did present. Those experts rely primarily on three pieces of evidence: (1) a study testing whether Celebrex could attenuate the tragic effects of Alzheimer's disease, known as the Alzheimer's 001 trial; (2) a so-called "meta-analysis" of Celebrex trials conducted by Dr. Madigan; and (3) with respect to Bextra, the CABG surgery trials.[14]

LEGAL STANDARD

In order for expert testimony to be admissible under Rule 702 and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), Plaintiffs must show that: (1) their proposed

[13] Plaintiffs thereby waived any ability to rely on the testimony of Drs. Bennett or Baruch for the purpose of opposing Defendants' motion. *See* Hr'g Tr. at 662-65.

[14] While Plaintiffs spent much of their briefing and time at the hearing identifying numerical differences in individual Bextra and Celebrex trials, they do not dispute that no Celebrex clinical trial before (or after) APC showed a statistically significant increase in the risk of heart attacks, strokes, or heart attacks and strokes combined, whether Celebrex was compared to placebo or other NSAIDs, nor do they dispute that no Bextra clinical trial other than the CABG trials showed a statistically significant increase in thrombotic risk. Using an endpoint that is consistent with the current boxed label for all NSAIDs, the lack of statistical significance prior to APC and outside the CABG trials is striking. *See* Defs.' Br., App., Figs. 4-6.

experts have the necessary qualifications in their fields of expertise; (2) their proposed testimony will assist the trier of fact; and (3) the experts' opinions are "based upon sufficient facts or data," are "the product of reliable principles and methods," and apply "the principles and methods reliably to the facts of the case." Fed. R. Evid. 702. In *Daubert*, the U.S. Supreme Court explained that district courts must perform a "gatekeeping" function to ensure that "any and all scientific testimony or evidence admitted is not only relevant, but reliable." 509 U.S. at 589.[15]

Even if a proposed expert witness possesses credentials to render certain expert opinions on one subject, the trial court should exclude any testimony that extends beyond the witness's demonstrated expertise. *See Quintanilla v. Komori Am. Corp.*, 2009 WL 320186, at *1 (2d Cir. Feb. 10, 2009); *McCullock v. H.B. Fuller Co.*, 981 F.2d 656, 657-58 (2d Cir. 1992). Moreover, a district court must "make certain that an expert . . . employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152 (1999). Numerous courts have recognized that expert testimony which incorporates only a portion of the available evidence, or "cherry-picks" the data that suits the expert's desired conclusion, is not sufficiently reliable and should be excluded under Rule 702 and *Daubert*.[16] Thus, where the "analytical gap" between the data and an

[15] *Daubert* identified several considerations as relevant in determining the reliability of proposed testimony: (1) whether the expert's technique or theory can be or has been tested; (2) whether the technique or theory has been subject to peer review and publication; (3) whether there are known or potential rates of error with regard to the specific techniques or theories when applied; and (4) whether the theory or approach has "general acceptance" in the relevant scientific community. *Id.* at 592-94. Courts – including this one – regularly exclude expert opinions that cannot be validated or that have unknown error rates. *See, e.g., Nimely v. City of N.Y.*, 414 F.3d 381, 399 (2d Cir. 2005); *In re Accutane Prods. Liab. Litig.*, 511 F. Supp. 2d 1288, 1296 (M.D. Fla. 2007); *In re Rezulin Prods. Liab. Litig.*, 369 F. Supp. 2d 398, 423 (S.D.N.Y. 2005).

[16] *See Lust v. Merrell Dow Pharms., Inc.*, 89 F.3d 594, 596 (9th Cir. 1996); *In re Rezulin*, 369 F. Supp. 2d at 437-38; *MTX Commc'ns Corp. v. LDDS/WorldCom, Inc.*, 132 F. Supp. 2d 289, 292 (S.D.N.Y. 2001).

expert's conclusions is "simply too great," the opinion is inadmissible. *See In re Bextra I*, 524 F. Supp. 2d at 1181 (*citing GE Co. v. Joiner*, 522 U.S. 136, 146 (1997)).

DISCUSSION

I. RELIABLE METHODS TO EVALUATE THROMBOTIC SAFETY OF NSAIDS.

Whether a researcher is designing a clinical trial, or a post hoc analysis, there are certain generally accepted principles that must be followed to ensure that the results are reliable. A researcher should: (1) systematically apply an "endpoint" or measure of the medical condition in question, which is accepted by medical doctors specializing in that condition; (2) reliably collect and diagnose medical outcomes of interest in a manner that follows accepted methods and criteria for diagnosing those outcomes; (3) use statistics to evaluate the probability that observed differences between the groups, if any, are due to the play of chance; and (4) use reliable methods for interpreting the results beyond the patients in the study. As demonstrated by the record citations below, these principles are undisputed by some or all of Plaintiffs' experts, whose own admissions related to these methodologies provide the Court with an adequate basis to judge whether Plaintiffs' experts deviated from reliable methodologies in this litigation. Accordingly, while helpful, the Court need not rely on the testimony of Defendants' experts.

A. To Evaluate the Thrombotic Risk of NSAIDs, the Scientific Community Has Analyzed Heart Attacks or Strokes Individually or Composite Endpoints That Include Heart Attacks and Strokes – But Not Composite Endpoints That Include Non-Thrombotic Events or Selectively Exclude Stroke.

1. Composite Endpoints Must Be Based on a Clinically Valid Rationale.

Prior to conducting a clinical trial or meta-analysis of trials, an investigator must pre-specify a research hypothesis that is valid according to medical doctors who specialize in the condition of interest. *See* Defs.' Ex. 198 at 11-12, 42 (EVALUATING CLINICAL RESEARCH ["EVALUATING"]); Defs.' Ex. 617 at 75, 85-86, 92-93 (Furberg Dep.); Defs.' Ex. 619 at 298

(Furberg & Morgan, STAT. MED. 1987;6:295-303 ["Furberg & Morgan"]). A physician with expertise in a medical condition or biological process of interest must specify the medical outcome of interest such as heart attacks (known as an endpoint), or a set of events or outcomes of interest, such as heart attacks, strokes, and other events precipitated by the same biological process (referred to as a composite endpoint). *See* Hr'g Tr. at 542. When using a composite endpoint to evaluate whether a medication affects a specific biological mechanism or process in the human body, the medical and scientific communities require the events included to be linked biologically by a common underlying disease mechanism, such as thrombosis in the arteries.[17]

A clinical trial or meta-analysis should have "*one* pre-specified primary hypothesis," Defs.' Ex. 198 at 43 (EVALUATING) – "the one the investigators are most interested in answering and . . . [that] is capable of being adequately answered." Defs.' Ex. 197 at 16 (FUNDAMENTALS OF CLINICAL TRIALS ["FUNDAMENTALS"]). Secondary endpoints should be interpreted cautiously, especially when the primary endpoint results are not statistically significant. *See* Defs.' Ex. 198 at 43 (EVALUATING). As with a primary endpoint, any secondary endpoint must

[17] *See* Hr'g Tr. at 532-34 (agreeing endpoints should relate to how a drug works in the body and that a clinician, not a statistician, must confirm that the endpoint is clinically relevant); Defs.' Ex. 198 at 90 (EVALUATING ["The components of the composite endpoint should make clinical sense."]); *id.* at 42 ("Red flags may include the use of unusual or illogical composites, *e.g.*, outcome measures that have uncertain clinical relevance."); Defs.' Ex. 1819 at 66 (Mem. to Arthritis Ad. Comm., May 15, 2009 [finding that "considering all cardiac serious adverse events as a single group of occurrences for purposes of evaluating safety is not informative"]). For example, when researchers study the anti-thrombotic (anti-clotting) effects of aspirin, they use a primary composite endpoint of biologically related, clot-induced events such as non-fatal heart attack, non-fatal stroke, and vascular death, known as the APTC endpoint. *See* Defs.' Ex. 620 at 82 (Antiplatelet Trialists' Collaboration, BRIT. MED. J., 1994;308:81-106). Plaintiffs' experts agree that the APTC endpoint is a commonly accepted way to measure thrombotic risk and have used it themselves. *See* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 205;111:249); Hr'g Tr. at 129-30, 329, 404; Defs.' Ex. 770 at 147 (Baruch Dep.); Defs.' Ex. 612 at 70-72, 263-64 (Bennett Dep., *In re Bextra*); Defs.' Ex. 617 at 34-35 (Furberg Dep.); Defs.' Ex. 198 at 89 (EVALUATING); Defs.' Ex. 671 (Folsom et al., ARCH. INTERN. MED. 2008;168(12):1333-39).

use biologically related events, and even then should be viewed as less definitive than primary hypotheses. *See* Defs.' Ex. 619 at 301 (Furberg & Morgan). Due to reliability concerns, secondary endpoints are used primarily to raise new hypotheses for further study. *See* Defs.' Ex. 197 at 17-18 (FUNDAMENTALS).

If a composite endpoint does not consist of individual conditions that are linked by a common biological process, statistical analyses that result from that endpoint are not a reliable basis upon which to evaluate whether a medication affects any particular biological process.[18] It is equally unreliable for a researcher to broaden or narrow a composite endpoint in order to achieve a desired statistical result,[19] which is why researchers always consult with medical experts and define the clinical question in the form of an endpoint before reviewing the event counts from the studies in question. *See* Defs.' Ex. 619 at 300 (Furberg & Morgan ["It is important that the data entered into a[n] overview be carefully considered by people with biomedical knowledge. Pooling should not be just a mechanical exercise with numbers!"]); Fed. Jud. Ctr., REF. MAN. ON SCI. EVID. 87 (2d ed. 2000) ("REF. MAN.").

2. Composite Endpoints Used to Evaluate the Thrombotic Safety of NSAIDs Consistently Focus on Thrombotic Events, Including Major Thrombotic Events Such as Heart Attack and Stroke.

Given the methodological issues related to endpoints, the parties spent a great deal of time discussing appropriate composite endpoints to evaluate the thrombotic safety of NSAIDs prior to APC. Fortunately, the Court need not weigh the testimony of the parties' experts or the arguments of counsel to arrive at an answer. Instead, the Court can look to objective and

[18] *See* Defs.' Ex. 619 at 301 (Furberg & Morgan); Defs.' Ex. 198 at 42 (EVALUATING); Defs.' Ex. 617 at 78-79 (Furberg Dep.).

[19] *See* Defs.' Ex. 198 at 11-12 (EVALUATING); *id.* at 39; Defs.' Ex. 2004 at I98 (Furberg et al., CIRCULATION 1983;67(suppl. I):I98-I101); Hr'g Tr. at 162.

undisputed sources to identify methodologically proper endpoints used by the medical and scientific communities at the time, including: (1) evidence regarding Plaintiffs' hypothesis for how Bextra and Celebrex work in the body; (2) peer-reviewed publications reporting the results of individual clinical trials and meta-analyses of clinical trials; and (3) endpoints used by the FDA. The consistency across these sources is telling, as every composite endpoint used to evaluate the thrombotic safety of NSAIDs prior to APC at a minimum included heart attacks and strokes – none selectively excluded heart attacks or strokes or lumped in cardiovascular conditions not precipitated by clots such as palpitations and arrhythmias.

First, Plaintiffs rely on the FitzGerald hypothesis – which they concede is based on an alleged clotting mechanism – to explain how selective COX-2 inhibitors increase the risk of heart attacks and strokes. *See* notes 11 & 12 *supra*. Dr. FitzGerald himself linked his hypothesis to thrombotic strokes and used the APTC endpoint with Dr. Furberg in their editorial. *See* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249). Dr. Bennett confirmed that researchers evaluating the hypothesis should measure heart attacks and strokes – testimony which stands unrebutted. *See* Defs.' Ex. 793 at 26 (Bennett Dep.). In Dr. Bennett's absence, Plaintiffs offer no biological basis to explain how selective COX-2 inhibitors could increase the risk of heart attacks, but not strokes, nor have they offered any other biological theory that was present in the literature before the announcement of the APC results.

Indeed, at the hearing, Dr. Furberg gave his clinical definition of a thrombotic event:

> Q. Your definition of thromboembolic includes clots that occur in the heart, brain and other parts of the body. Is that correct?
>
> A. Well, that is what I said. The thrombotic events are those that are linked to a clot, and the clot can happen anywhere in the body. Typically in my field, the medical field, we consider those in the heart because heart attacks, we consider those to the brain. We consider those in the lungs and those peripherally, and those are all thrombotic or thromboembolic events. That is the clear definition.

Hr'g Tr. at 235.[20] Similarly, Dr. Furberg testified that "you have to take into account the mechanism of action" in formulating an endpoint, that a general cardiovascular endpoint is "too broad" because some cardiovascular conditions are not affected by selective COX-2 inhibitors, and that researchers should include stroke to measure the thrombotic risk described in the Celebrex boxed warning. Hr'g Tr. at 241 (citing Furberg Dep. at 428-29). In fact, in an editorial Drs. FitzGerald and Furberg published regarding Bextra, they did just that, selecting a composite of heart attacks and strokes as their endpoint to evaluate thrombotic risk in the CABG trials because those events were the "most relevant" at the time. Hr'g Tr. at 233-34 (citing Furberg Dep. at 212); *see* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249).[21]

In light of the relevant biological hypothesis at the time, it is not surprising that every peer-reviewed composite endpoint intended to measure the thrombotic safety of selective COX-2 inhibitors before APC included heart attack and stroke events at a minimum. For example, the APPROVe and CABG-2 trials both prospectively defined thrombotic endpoints to include heart attacks and strokes in the same analysis.[22] Dr. Kearney used the APTC endpoint as her primary endpoint in her meta-analysis, and then examined its individual components, *see* Defs.' Ex. 784 at 1303 (Kearney), an endpoint methodology Dr. Furberg called "as good as it gets." Hr'g Tr. at

[20] *See id.* at 244-45 (Dr. Furberg agreeing with the inclusion of stroke in the boxed warning); Defs.' Ex. 616 at 397 (Furberg Dep., *Haslam v. Pfizer* [testifying that heart attacks and stroke "go hand in hand" and that he cannot distinguish Bextra's heart attack risk from its stroke risk]).

[21] Drs. Bennett and Furberg also contributed to the AHA statements, which used the APTC endpoint and included stroke. *See* Defs.' Ex. 740 at 1715 (Bennett et al., CIRCULATION 2005;111:1713-16 ["2005 AHA statement"]); Defs.' Ex. 682 at 1640 (2007 AHA statement).

[22] *See* Defs.' Ex. 638 at 1094 (Bresalier et al., N. ENGL. J. MED. 2005;352:1092-1102 [APPROVe]); Defs.' Ex. 696 at 1082 (Nussmeier [CABG-2]); *see also* Hr'g Tr. at 235-36, 241-42 (Dr. Furberg testifying the CABG-2 endpoint was appropriate); Defs.' Ex. 617 at 213, 429 (Furberg Dep. [same]). The FDA also used the APTC endpoint to evaluate APC and PreSAP. Defs.' Ex. 664 at 4 (FDA Decision Mem.).

279-80. Dr. Madigan included both heart attack and stroke in the primary endpoint for his Vioxx litigation analysis, even though he has no reason to believe Vioxx and Celebrex have different thrombotic effects. *See* Hr'g Tr. at 545, 548; Defs.' Ex. 324 at 14, 33 (App. A) (Madigan Vioxx Rep.). Indeed, *Plaintiffs have not identified a single trial evaluating the thrombotic safety of NSAIDs that combined heart attacks with other events, but not stroke, before December 2004*.

Finally, the FDA also focused on thrombotic events and included strokes in its analyses of NSAIDs. As early as 2001, the FDA revised the labels for both Celebrex and Vioxx to reflect the thrombotic results – including stroke – from the respective CLASS and VIGOR trials, not some composite grouping that lacked a common biological mechanism or some endpoint that excluded stroke. In its decision memorandum in 2005, the FDA used the APTC composite endpoint, which includes strokes. *See* Defs.' Ex. 664 at 4 (FDA Decision Mem. [calling APTC "widely accepted"). Even after APC, the FDA continued to use a thrombotic endpoint that included stroke; it did not add a warning regarding all cardiovascular conditions of any kind, it did not warn only of conditions that occur in the vicinity of the heart, and it did not omit stroke.

This consistent body of evidence renders unpersuasive Plaintiffs' counsels' attempts at the hearing to group different types of cardiovascular conditions as ones that occur "in the heart" or "in the head." Such an approach has no clinical validity and is contrary to the testimony of Plaintiffs' own medical experts. Thus, a reliable analysis of the thrombotic safety of Bextra or Celebrex must focus on thrombotic events, and any composite thrombotic endpoint must include the major thrombotic events such as heart attacks and strokes.

B. To Evaluate the Thrombotic Safety of NSAIDs, Researchers Must Collect and Diagnose Conditions Reliably.

During a clinical trial, patient events are reported by physicians who serve as field investigators at various study locations. As a result, events may be reported in different ways by

different investigators, and even different ways by the same doctor at different times. *See* Defs' Ex. 616 at 777 (Furberg Dep., *Haslam v. Pfizer*); Defs.' Ex. 617 at 427-28 (Furberg Dep.). To eliminate these inconsistencies, many trials have "adjudication committees," which are groups of medical experts who specialize in the outcomes of interest and review all available information about each patient without knowing whether the patients are taking the medication being tested or placebo pills (a process known as "blinding") using pre-specified, written procedures and diagnostic criteria to ensure that all events are defined, counted, and classified consistently.[23]

In order for a committee of medical experts to conduct its blinded review and decide whether patients experienced conditions that fit a defined endpoint, researchers first must collect all the relevant recorded events and provide the committee with all available medical information for each patient in question. These committees use the detailed, individual patient narratives from each study, which usually are the most reliable source of original patient information, to ensure that their diagnoses are based on the most complete and accurate information available. *See* Defs.' Ex. 2014 at 6-7, 8 (2005 FDA Reviewer Guidance); Defs.' Ex. 197 at 311 (FUNDAMENTALS); Defs.' Ex. 622 at 62-63 (Madigan Dep.). Short event descriptions in computer data files (known as "SAS" files) may use codes and short terms that do not properly describe what actually happened to patients in a trial. *See* Defs.' Ex. 2014 at 8 (2005 FDA Reviewer Guidance); Hr'g Tr. at 680-81; Defs.' Ex. 2023 at 18-19 (Dr. Packer Demonstratives).

The committee of medical experts also must define expressly the criteria for diagnosing each component of an endpoint before the review commences. Board certified cardiologists

[23] *See* Defs.' Ex. 2014 at 8 (2005 FDA Reviewer Guidance) Defs.' Ex. 617 at 105 (Furberg Dep.); Ex. 616 at 232-33, 769-70, 781 (Furberg Dep., *Haslam v. Pfizer*); Defs.' Ex. 770 at 80-82, 96-97 (Baruch Dep.); Defs.' Ex. 621 at 129, 230-31 (Kronmal Dep.); Defs.' Ex. 622 at 62-63 (Madigan Dep.); Defs.' Defs.' Ex. 198 at 51 (EVALUATING).

should be a part of the diagnosis process when thrombotic events are at issue.[24] When more than one reviewer classifies events, there should be consensus on each event, with disagreements resolved by a third reviewer using a pre-specified process. *See* Defs.' Ex. 617 at 388-91 (Furberg Dep.); Defs.' Ex. 616 at 770 (Furberg Dep., *Haslam v. Pfizer*). It is inappropriate to arbitrarily pick one classification over another. *See* Defs.' Ex. 617 at 417-18 (Furberg Dep.).

Several examples in the NSAID literature demonstrate the application of these principles, including APC, APPROVe, and CABG-2. In each of those trials: (1) the researchers defined their thrombotic endpoint, which included stroke, before they knew the results of the trial; (2) a well-qualified cardiovascular safety committee defined in advance the criteria used to evaluate each component of the primary composite endpoint; and (3) a blinded adjudication committee, which included experienced cardiologists, reviewed all available narratives and other medical information for each patient and then decided whether a reported event fit the pre-specified criteria for any component of the primary composite endpoint.[25]

C. For Common Injuries Like Heart Attacks and Strokes, Researchers Require Statistical Significance to Reliably Assess Whether There Is Increased Risk.

Statistical significance thresholds (expressed as *p*-values and/or confidence intervals) "permit an assessment of whether the results of a study are likely to represent a true association

[24] *See* Hr'g Tr. at 870; Defs.' Ex. 793 at 69-70, 72-73 (Bennett Dep.). A doctor who does not specialize in treating patients in the relevant field is not qualified to determine or classify the causes of events. *See Gayton v. McCoy*, 521 F. Supp. 2d 841, 847-48 (C.D. Ill. 2007); *Abdoush v. Jackson*, 2007 WL 4557711 at *3-5 (E.D. Mich. Dec. 19, 2007).

[25] *See* Defs.' Ex. 658 at 1072-73 (Solomon S. et al., N. ENGL. J. MED. 2005;352:1071-80); Defs.' Ex. 638 at 1094-95 (Bresalier et al, N. ENGL. J. MED. 2005;352:1092-1102); Defs.' Ex. 696 at 1082-83 (Nussmeier). The Pfizer-sponsored and peer-reviewed White meta-analysis, which Dr. Madigan ignored in favor of an unblinded meta-analysis not performed by cardiologists or validated by peer review, also relied on an adjudication committee with "extensive expertise" in endpoint assessment that performed its blinded review based on pre-specified diagnostic criteria. Defs.' Ex. 1815 at 91, 93 (White et al., AM. J. CARDIOL. 2007;296:91-98).

or random error." REF. MAN. at 354. In other words, a *p*-value functions as an error rate by indicating the probability that the observed difference will arise by chance alone if the trial is repeated. *See id.* at 357; Defs.' Ex. 198 at 107, 141 (EVALUATING). Where a single statistical test is pre-specified, a *p*-value of 0.05 means that there is a 5% chance that the result was due to chance alone (*i.e.*, a 5% error rate). *See* REF. MAN. at 357-58. If the observed *p*-value is less than the pre-specified *p*-value threshold, then the result is statistically significant. *See id.* at 396. Where a single statistical test is pre-specified, the most conventional *p*-value is < 0.05. *See* Defs.' Ex. 611 at 6 (Kronmal Rep.).[26]

1. The Medical and Scientific Communities, Courts Evaluating Securities Claims, and Courts Applying *Daubert* All Require Statistical Significance as a Prerequisite for Reliability.

Because of the importance of statistical significance in evaluating common events such as heart attacks and strokes, the FDA and the rest of the scientific and medical communities always have relied on statistical significance when assessing the cardiovascular safety of NSAIDs.[27] Indeed, Plaintiffs have not identified a single peer-reviewed study discussing the cardiovascular safety of NSAIDs that did not base conclusions on the statistical significance of the results. Similarly, outside the context of NSAIDs, the FDA universally requires evidence of a

[26] Statistical significance does not establish a causal relationship, only that two events may be associated. *See* Defs.' Ex. 617 at 136-37 (Furberg Dep.). Researchers must examine the totality of evidence, because statistically significant results that are replicated in and consistent across studies provide confidence that observed differences are real. *See* Hr'g Tr. at 218; Defs.' Ex. 612 at 76, 204, 379 (Bennett Dep., *In re Bextra*); Defs.' Ex. 617 at 62-63 (Furberg Dep.); *see also* REF. MAN. at 102 (noting that reproducibility is a hallmark of reliability). Researchers also must offer a valid biological mechanism for how a drug could cause an alleged effect. *See* Defs.' Ex. 612 at 539 (Bennett Dep., *In re Bextra*).

[27] *See, e.g.*, Defs.' Ex. 664 at 5 (FDA Decision Mem. [noting that the results of the Alzheimer's 001 study "did not demonstrate a significantly increased risk of serious adverse CV events" and discussing APPROVe and VIGOR trials]); Defs.' Ex. 658 at 1073 (Solomon S. et. al., N. ENGL. J. MED. 2005;352:1071-80 [APC]); Defs.' Ex. 683 at 1630 (McGettigan).

statistically significant association between an adverse event and a medication where the alleged injury occurs commonly in the general population – such as heart attacks and strokes – in order to distinguish a potential heart attack and stroke effect of a medication from the many other well-known causes of these conditions (*e.g.*, high cholesterol, smoking, and family history). It is only where an adverse event is idiosyncratic and rarely occurs in the absence of a medication that the FDA will make safety decisions without statistically significant evidence. *See* Defs.' Exs. 825 – 1026; *see also* Defs.' Reply Br., App., Fig. 13 (describing "Type A" and "Type B" reactions).[28]

The Second Circuit likewise requires plaintiffs asserting federal securities law claims based on a failure to disclose adverse events allegedly associated with a medication to plead and prove that there was statistically significant evidence of a causal relationship between the medication and the alleged harm. *See Honeyman v. Hoyt (In re Carter-Wallace, Inc. Sec. Litig.)*, 150 F.3d 153, 157 (2d Cir. 1998) ("*Carter-Wallace I*") ("Drug companies need not disclose isolated reports of illnesses suffered by users of their drugs until those reports provide statistically significant evidence that the ill effects may be caused by – rather than randomly associated with – use of the drugs and are sufficiently serious and frequent to affect future earnings."); *see also Honeyman v. Hoyt (In re Carter-Wallace, Inc. Sec. Litig.)*, 220 F.3d 36, 42 (2d Cir. 2000) ("*Carter-Wallace II*").[29] This rule makes abundant sense, because requiring

[28] Plaintiffs also do not refute case law establishing that the FDA's precautionary public health determinations regarding labeling do not provide a sufficiently reliable basis for an expert opinion that a medication is associated with an adverse event. *See Rider v. Sandoz Pharms. Corp.*, 295 F.3d 1194, 1201 (11th Cir. 2002); *Glastetter v. Novartis Pharms. Corp.*, 252 F.3d 986, 991 (8th Cir. 2001); *Lopez v. Wyeth-Ayerst Labs., Inc.*, 1998 WL 81296, at *2 (9th Cir. Feb. 25, 1998); *Newton v. Roche Labs., Inc.*, 243 F. Supp. 2d 672, 683 (W.D. Tex. 2002); *Nelson v. Am. Home Prods. Corp.*, 92 F. Supp. 2d 954, 958 (W.D. Mo. 2000).

[29] The principle articulated in *Carter-Wallace I* and *II* repeatedly has been affirmed by the Second Circuit, as recently as August 2009. *See Avon Pension Fund v. GlaxoSmithKline PLC*, 2009 WL 2591173, at *1 (2d Cir. Aug. 24, 2009) ("Reports or test results must yield reliable

pharmaceutical companies to disclose the results of every clinical trial, even where such trials do not yield statistically significant evidence of a true causal effect, would "risk 'bury[ing] the shareholders in an avalanche of trivial information.'"[30]

Similarly, the great majority of courts evaluating the admissibility of causation opinions under *Daubert* require statistical significance because it "bears heavily on . . . reliability for evidential purposes."[31] The few contrary cases Plaintiffs cite either involved very rare injuries, dismissed plaintiffs' claims, or involved questions other than reliability under *Daubert*.[32] Indeed, Plaintiffs have not cited a single *Daubert* decision in which a court abandoned statistical significance in the context of common injuries like heart attacks and strokes.

---

evidence of a drug's adverse effect to give rise to a duty of manufacturers to disclose those results to potential investors. While the complaint conclusorily alleges that the results of the meta-analyses 'showed an estimate' of an 'increased risk of heart attack,' it pleads no facts indicating that the test results were even statistically significant.") (internal citations omitted); *see also State Univs. Retirement Sys. of Ill. v. AstraZeneca PLC*, 2009 WL 1796534, at *2 (2d Cir. June 25, 2009); *Masters v. GlaxoSmithKline*, 271 Fed. Appx. 46, 51 (2d Cir. 2008).

[30] *Borochoff v. GlaxoSmithKline PLC*, 2008 WL 2073421, at *7 (S.D.N.Y. May 9, 2008) (*quoting San Leandro Emergency Med. Group Profit Sharing Plan v. Philip Morris, Co.*, 75 F.3d 801, 810 (2d Cir. 1996)), *aff'd sub nom. Avon Pension Fund v. GlaxoSmithKline PLC*, 2009 WL 2591173 (2d Cir. Aug. 24, 2009) (Summary Order).

[31] *DeLuca v. Merrell Dow Pharms., Inc.*, 791 F. Supp. 1042, 1057 (D.N.J. 1992) (citations omitted), *aff'd without opinion*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 510 U.S. 1044 (1994); *see Joiner*, 522 U.S. at 146; *Allen v. Penn. Eng'g Corp.*, 102 F.3d 194, 197 (5th Cir. 1996); *Smith v. Wyeth-Ayerst Labs. Co.*, 278 F. Supp. 2d 684, 691 (W.D.N.C. 2003); *Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 533 (W.D. Pa. 2003); *In re Norplant Contraceptive Prods. Liab. Litig.*, 215 F. Supp. 2d 795, 831 (E.D. Tex. 2002); *Caraker v. Sandoz Pharms. Corp.*, 188 F. Supp. 2d 1026, 1034 (S.D. Ill. 2001); *Jones v. U.S.* 933 F. Supp. 894, 898-900 (N.D. Cal. 1996); *Thomas v. Hoffman-LaRoche, Inc.*, 731 F. Supp. 224, 228 (N.D. Miss. 1989).

[32] Plaintiffs' citation to the *Seroquel* and *Neurontin* decisions are unavailing, as the experts there relied on statistically significant evidence *in addition to* non-significant evidence, or epidemiological evidence was not reasonably available. *See In re Seroquel Prods. Liab. Litig.*, No. 6: 06-MD-1769, slip. op., at *22 (M.D. Fla. June 18, 2009); *In re Neurontin Mktg., Sales Practices, & Prods. Liab. Litig.*, 612 F. Supp. 2d 116, 140-41 (D. Mass. 2009).

2. Opinions Based on "Signals" and "Trends" Are Not Reliable.

Even though statistical significance is the recognized, established, and reliable basis for assessing data for injuries like heart attacks and strokes, Plaintiffs' experts base their opinions here on "trends" and "signals." Those methodologies are unreliable for several reasons. First, there are no objective criteria for assessing whether or not a "trend" or "signal" exists. *See* Hr'g Tr. at 222; Defs.' Ex. 617 at 70, 244 (Furberg Dep.); Defs.' Ex. 793 at 50 (Bennett Dep.). In fact, Plaintiffs' experts admit that "trends" are not reliable to assess whether a medicine increases or decreases risk, but are simply hints of issues that may require further study. *See* Hr'g Tr. at 121-22, 215-17; Defs.' Ex. 793 at 55 (Bennett Dep.). Because "trends" and "signals" are not reliable to establish a drug effect, they are "often wrong." Defs.' Ex. 783 at 51 (Bennett Dep.); *see* Hr'g Tr. at 225-29 (discussing instances where conclusions based on signals were not replicated in subsequent research). As further evidence of the inherent lack of reliability in using "signals" to detect true associations, Dr. Kronmal has gone so far as to say that one could label *any* difference in the number of events as a signal. *See* Defs.' Ex. 621 at 186 (Kronmal Dep.).

Plaintiffs' own experts recognize that a *p*-value functions as an error rate by indicating the probability that the observed difference will arise by chance alone if the trial is repeated. *See* Defs.' Ex. 197 at 335 (FUNDAMENTALS); Defs.' Ex. 621 at 147 (Kronmal Dep.); Defs.' Ex. 611 at 6 (Kronmal Rep.). Plaintiffs' experts also repeatedly recognize the importance of statistical significance as a prerequisite for reaching reliable conclusions about the safety of a medication. *See* Hr'g Tr. at 214 (Dr. Furberg); Defs.' Ex. 793 at 54-55, 104-06 (Bennett Dep.); Defs.' Ex. 617 at 134, 142 (Furberg Dep.); Defs.' Ex. 622 at 136-37, 158 (Madigan Dep.). Plaintiffs' reliance here on "scientific significance" – a concept they attempt to distinguish from *statistical* significance – is meaningless, as *"[a]n important objective measure of scientific significance is 'statistical significance,'* which usually requires that there is no more than a 5 percent probability

that the scientist's findings are the results of chance." Defs.' Ex. 816 at 87 n.3 (Mark G. Haug, *Minimizing Uncertainty in Scientific Evidence* in SCIENTIFIC EVIDENCE REVIEW: CURRENT ISSUES AT THE CROSSROADS OF SCIENCE, TECHNOLOGY AND THE LAW (2006) (emphasis added)).

3. Statistical Significance Thresholds Must Be Adjusted to Reflect the Use of Multiple, Post Hoc Comparisons of Secondary Endpoints, Particularly in Pooled Analyses of Clinical Trials.

Researchers do not always consider statistical significance in the context of a primary hypothesis in a single study. At times, researchers conducting post hoc analyses may compare dozens of variables within an individual trial or across multiple trials. If a post hoc researcher compares enough variables between two treatment groups, some of the comparisons may appear to be statistically significant by chance alone when in fact no difference exists, which is why researchers call such results "nominally" significant.[33] Under such circumstances, using a conventional *p*-value threshold is unreliable because the true error rate is much higher.[34]

---

[33] *See* REF. MAN. at 127; Hr'g Tr. at 281-82; Defs.' Ex. 1819 at 66 (Mem. to Arthritis Ad. Comm., May 15, 2009); Defs.' Ex. 197 at 124, 339 (FUNDAMENTALS); Defs.' Ex. 198 at 111 (EVALUATING). For example, if one were to compare a medication against placebo across 600 variables and use a *p*-value of 0.05 for each of the comparisons, one would expect 30 of the 600 comparisons to appear nominally statistically significant by chance alone, even if there were no relationship between the medication and the variables being tested. *See* REF. MAN. at 166. To demonstrate this statistical artifact, known as a "multiple comparison" problem, researchers have found nominally statistically significant associations between patients' astrological signs and reasons for hospitalization, despite the lack of any causal relationship. *See* Hr'g Tr. at 747-48; Defs.' Ex. 2017 at 964 (Austin et al., J. CLIN. EPID. 2006;59:964-69).

[34] *See* REF. MAN. at 128 ("In these situations, courts should not be overly impressed with claims that estimates are significant."); Defs.' Ex. 619 at 296 (Furberg & Morgan ["[C]onclusions based on these results are unreliable owing to the numerous tests (reported or unreported) that were performed."]); *see also* Defs.' Ex. 197 at 311-12 (FUNDAMENTALS); Defs.' Ex. 198 at 111 (EVALUATING); Defs.' Ex. 615 at 70-71 (Furberg Dep., *Valenzuela v. Warner-Lambert Co.*); Defs.' Ex. 617 at 75 (Furberg Dep.); Defs.' Ex. 2004 at I98 (Furberg et al., CIRCULATION 1983;67(suppl. I):I98-I101); Defs.' Ex. 2006 at 2 (Beta-Blocker Heart Attack Trial Research Group, J. AM. COLL. CARDIOL. 1986;7:1-8).

Consequently, where researchers compare dozens of variables, the medical and scientific communities either discount the statistical results or require a downward adjustment of the statistical significance threshold.[35] Such adjustments are particularly appropriate for results involving secondary endpoints, which are less reliable than primary endpoints.[36] Without such an adjustment, there is a greatly increased probability of an erroneous conclusion. *See* Hr'g Tr. at 537-38 (admitting that failing to adjust increases the risk of Type I errors, *i.e.*, falsely claiming an association when one does not exist); Defs.' Ex. 617 at 77-78 (Furberg Dep. ["[T]here's no way of coming up with a proper significance level *p*-value for post hoc analyses."]).

Indeed, researchers studying the thrombotic effects of NSAIDs – particularly those performing meta-analyses – regularly adjust the *p*-value threshold to account for the existence of multiple comparisons. For example, the Kearney meta-analysis – on which several of Plaintiffs' experts rely – pre-specified a *p*-value of 0.01 for her secondary endpoints. *See* Defs.' Ex. 784 at 1303 (Kearney [specifying a 99% confidence interval to allow for the multiplicity of analyses]); Hr'g Tr. at 281 (Dr. Furberg applauding Kearney for using 99% confidence intervals and noting that she "must have read our book"). While some of Plaintiffs' experts claim for litigation purposes that such adjustments should not be made in the context of safety, they did not identify

[35] *See* Defs.' Ex. 198 at 123 (EVALUATING ["Post hoc analyses of data derived from clinical trials designed to answer other research questions are perhaps the least reliable."]); Hr'g Tr. at 261-62 (testifying that a nominal *p*-value of 0.024 could be due to chance due to the multiplicity of comparisons and noting that researchers would have to replicate that result before it could be considered reliable); *see also id.* at 281-82; Defs.' Ex. 197 at 124, 305, 339 (FUNDAMENTALS); Defs.' Ex. 617 at 115-16 (Furberg Dep.).

[36] *See* Defs.' Ex. 197 at 307-08 (FUNDAMENTALS [advising caution in claiming significant results for comparisons other than the primary outcome]); Defs.' Ex. 198 at 43 (EVALUATING ["When the results for the primary hypothesis are not significant, be cautious if the focus is shifted to a secondary endpoint, or one that is defined post hoc."]); *see also* Defs.' Ex. 789 at 262 (Boissell et al., CONTROLLED CLIN. TRIALS 1989;10:254-281); Defs.' Ex. 617 at 77-78 (Furberg Dep.); Defs.' Ex. 619 at 301 (Furberg & Morgan).

any published literature advocating that approach, let alone endorsing a Madigan-like analysis that was based on more than 600 *p*-value calculations.[37]

4. A *P*-Value Only Applies to the Specific Population Studied.

A statistically significant *p*-value from one study is not a reliable basis to support an increased risk opinion related to patients who receive different medications and different dosages under different clinical conditions.[38] In particular, it is inappropriate to apply *p*-values from studies involving one form of a medication to patients taking another form – for example, from an intravenous administration to oral pills – because each form may exert different effects.[39]

II. PLAINTIFFS' EXPERTS' ALZHEIMER'S 001 OPINIONS RELY ON AN UNPRECEDENTED ENDPOINT WHICH INCLUDES NON-THROMBOTIC EVENTS, MAKING IT UNRELIABLE FOR ASSESSING THROMBOTIC RISK.

Because no Celebrex trial before APC showed a statistically significant increase in thrombotic risk, Plaintiffs' experts rely heavily on the Phase 2 Alzheimer's 001 trial, a high-dose, experimental trial in 425 patients designed to evaluate whether Celebrex could attenuate

---

[37] Indeed, when asked at the hearing about the hypotension safety data in the CABG surgery trials, Dr. Furberg testified that those results may not be reliable because it is common to find nominally statistically significant results when one performs multiple comparisons. *See* Hr'g Tr. at 261. Similarly, he asserted that the ADAPT safety results showing that naproxen (Aleve) *increased* the risk of thrombotic events compared to placebo were unreliable due to the use of multiple comparisons. *See* Defs.' Ex. 617 at 115-16, 331-32 (Furberg Dep.).

[38] *See* Defs.' Ex. 626 at 88 (Furberg, POSTGRADUATE MEDICINE: A QUARTER CENTURY OF BETA BLOCKADE (1988) ["Extrapolation to subsets of patients that would not have qualified for enrollment into the trial represents a leap of faith."]); Defs.' Ex. 627 at 304-05 (Jewell Dep., *In re Bextra* ["To extrapolate to populations beyond which the sample was drawn requires something other than statistics."]); David L. Eaton, 12 J. L. & POL'Y 5, 11 (2003) (calling dose "the single most important factor to consider in evaluating whether an alleged exposure caused a specific adverse effect"); *see also* Defs.' Ex. 616 at 142-43, 151-54, 166-67, 282-83, 444-46, 456 (Furberg Dep., *Haslam v. Pfizer*); Defs.' Ex. 612 at 220-21, 279-80, 328-29 (Bennett Dep., *In re Bextra*); Defs.' Ex. 632 at 168-69 (Ad. Comm. Tr., Feb. 18, 2005); Defs.' Ex. 197 at 37, 341 (FUNDAMENTALS); Defs.' Ex. 198 at 116, 131 (EVALUATING).

[39] *See* Hr'g. Tr. at 265-68; Defs.' Ex. 197 at 181-82 (FUNDAMENTALS); Defs.' Ex. 623 at IV-18 (Furberg, CLIN. CARDIOL. 2000;23;7 Suppl. 4:IV15-19); Defs.' Ex. 621 at 318 (Kronmal Dep.).

the progression of Alzheimer's disease. In their post-hoc analysis of this trial, Plaintiffs' experts rely on composite measures that either include non-thrombotic events like heart palpitations or selectively exclude major thrombotic events such as stroke, or both. Outside of this litigation, no such composite endpoints ever have been used to evaluate the thrombotic safety of a medication, and Plaintiffs' own experts concede that their endpoints do not evaluate the heart attack or stroke risk of Celebrex. Because these composite endpoints are not recognized measures of thrombotic risk, they cannot provide a reliable basis to assess thrombotic risk, and opinions based on them therefore fail to satisfy *Daubert*.

A. The FDA, the Investigators, and Plaintiffs' Experts Agree the Trial Does Not Demonstrate a Statistically Significant Increase in Thrombotic Risk.

The FDA, the independent researchers who conducted the Alzheimer's 001 trial, and Plaintiffs' own experts agree that the trial does not provide reliable evidence of an increased thrombotic risk. At the Advisory Committee meeting, Dr. James Witter of the FDA noted that the results were "difficult to interpret . . . because of the small sample size which made relative risk and odds ratios unreliable." Defs.' Ex. 632 at 398 (Ad. Comm. Tr., Feb. 16, 2005 [calling it "difficult to know how to generalize these results"]). He also observed that the Celebrex patients in the study had baseline "imbalances in terms of hypertension, diabetes, those that had bypass surgery, those that had history of ischemia and those that had history of coronary-artery disease" compared to the placebo group. *Id.*; *see* Defs.' Ex. 667 at 36 (Presentation by Dr. Witter, Feb. 16, 2005); Defs.' Ex. 673 at 12-13, 17, 19 (Soininen at al., DEMENT. GERIATR. COGN. 2007;23:8-21 ["Soininen"]). In other words, the Celebrex group was materially sicker than the placebo group *before* either group took its first dose of study medication. Following the Advisory Committee meeting, the FDA specifically found no statistically significant increase in the risk of thrombotic events in the Alzheimer's 001 trial. *See* Defs.' Ex. 664 at 5 (FDA Decision Mem.).

The FDA's determination was consistent with the peer-reviewed, published conclusions of the investigators who conducted the trial, who noted that the results were driven not by thrombotic events such as heart attacks and strokes, but instead by non-thrombotic events unrelated to Plaintiffs' proffered FitzGerald hypothesis and the public health debate that led to the boxed label for all NSAIDs. *See* Defs.' Ex. 673 at 18-19 (Soininen ["These differences were primarily driven by the individual terms cardiac failure, fibrillation atrial, and angina pectoris."]). The investigators also confirmed the baseline imbalances noted by Dr. Witter. *See id.* at 13. Plaintiffs' own experts Drs. Bennett, Furberg, and Kronmal agree that the Alzheimer's 001 trial does not show an increased risk of thrombotic events such as heart attacks.[40]

B. Plaintiffs' Experts Rely on Unprecedented Composites of Events That Are Unreliable for Assessing Thrombotic Risk.

Because the Alzheimer's 001 trial did not show an increased thrombotic risk, Plaintiffs' experts ignore the baseline cardiovascular imbalances in the trial and fashion their cardiovascular risk argument around collections of "cardiovascular" – not thrombotic – adverse events, most of which never have been linked to the use of Celebrex or any other NSAID. These analyses are not reliable (or even relevant) because they bear no relation to any accepted measure of thrombotic risk, the current boxed label, the FitzGerald hypothesis, or Plaintiffs' complaint.

Dr. Furberg's endpoint combines the "cardiovascular" events of stroke, heart failure, pulmonary edema, myocardial infarction, angina pectoris and atrial fibrillation. Defs.' Ex. 610 at 18 (Furberg Rep.); Defs.' Ex. 617 at 148 (Furberg Dep.). He admits, though, that his composite

[40] *See* Hr'g Tr. at 411-12 (Dr. Kronmal admitting the numbers of heart attacks were "way too small to reach any strong conclusions"); Defs.' Ex. 621 at 298 (Kronmal Dep. [same]); Defs.' Ex. 793 at 85 (Bennett Dep. [calling it "impossible" to interpret the heart attacks in the trial]); Hr'g Tr. at 179 (Dr. Furberg identifying APC as the first trial showing a significant increase in thrombotic risk); *see also* Defs.' Ex. 793 at 79 (Bennett Dep.); Defs.' Ex. 617 at 162-63 (Furberg Dep.); Defs.' Ex. 621 at 294 (Kronmal Dep.); Defs.' Ex. 611 at 22-23 (Kronmal Rep.).

endpoint is not intended to measure thrombotic risk, and Plaintiffs' other experts agree. *See* Hr'g Tr. at 246; Defs.' Ex. 617 at 28 (Furberg Dep. [angina]); Defs.' Ex. 770 at 240 (Baruch Dep. [atrial fibrillation, angina]); Defs.' Ex. 793 at 26 (Bennett Dep. [atrial fibrillation]). Dr. Furberg even downplayed the statistical significance of his unique endpoint under oath, calling the result "nominally" significant and adding, "It's more of a signal to say – to flag that something may not be right here." Defs.' Ex. 617 at 163 (Furberg Dep.). Dr. Kronmal's endpoint consisted of events falling within certain "heart-related" body system terms from a medical dictionary. *See* Defs.' Ex. 611 at 22 (Kronmal Rep.); Defs.' Ex. 57 at 6 (PhRMA web site). These terms also do not represent a thrombotic endpoint, but rather various cardiovascular events generally, as Dr. Kronmal admits. *See* Hr'g Tr. at 404-05; Defs.' Ex. 793 at 79 (Bennett Dep.).

In their entire careers in clinical research, neither Dr. Furberg nor Dr. Kronmal ever has used these combinations of events as composite endpoints. *See* Hr'g Tr. at 248-50 (Dr. Furberg admitting further that he did not pre-specify his Alzheimer's 001 endpoint and instead observed trends in certain events and added them together); Defs.' Ex. 617 at 154 (Furberg Dep.); Defs.' Ex. 621 at 249-50 (Kronmal Dep.). In fact, there is no evidence in the record that such endpoints ever have been used in a clinical trial or in the medical literature *in any context*. Notably, outside of this litigation, Dr. Furberg has used the APTC endpoint to assess thrombotic risk, including for selective COX-2 inhibitors. *See* Defs.' Ex. 617 at 34-35, 131, 158 (Furberg Dep.). Indeed, Drs. Furberg and FitzGerald used a thrombotic endpoint that included heart attacks and strokes to evaluate thrombotic risk in the CABG trials – but did not include any of the non-thrombotic events Dr. Furberg uses in his Alzheimer's 001 endpoint for this litigation.[41]

[41] *See* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249). In addition to using an inappropriate combined endpoint, Dr. Furberg ignored a material portion of the data associated

III. ALL THE ITERATIONS OF DR. MADIGAN'S META-ANALYSES EMPLOY CLINICALLY INVALID ENDPOINTS, A FLAWED DATA COLLECTION METHOD, AND AN UNDOCUMENTED, UNPRECEDENTED PROCEDURE TO DIAGNOSE THE CONDITIONS OF THE PATIENTS IN THE TRIALS.

Because the Alzheimer's 001 trial did not show a statistically significant increase in thrombotic risk, Plaintiffs also rely on Dr. Madigan's meta-analyses, which are unreliable for a number of reasons. Plaintiffs' own experts admit that post hoc meta-analyses – particularly those analyzing secondary endpoints among hundreds of comparisons – are not definitive. Even with those limitations, Dr. Madigan's analyses only purport to show statistical significance for two of his four endpoint hypotheses, and only at the highest Celebrex doses, based primarily on a handful of events from Alzheimer's 001. The portions of statistical tests on which Dr. Madigan now relies to support his opinions are based on clinically invalid endpoints that never have been used in the peer-reviewed clinical trial literature to evaluate the thrombotic safety of NSAIDs and on a flawed data collection and classification procedure. As a result, the various iterations of Dr. Madigan's meta-analyses are unreliable and inadmissible under *Daubert* and Rule 702.

A. Plaintiffs' Own Experts Warn That Meta-Analyses Are Not Definitive, and Dr. Madigan's Constantly Evolving Analyses Warrant Additional Scrutiny.

According to Plaintiffs' experts, meta-analyses should be a "secondary consideration" used primarily to generate hypotheses for future study.[42] Because meta-analyses are post hoc, an investigator's bias may affect the results. *See* Defs. Ex. 617 at 95-96 (Furberg Dep.). "One needs to be particularly cautious about pooling data not directly related to the main objective of

---

with the Alzheimer's 001 study. *See* Defs.' Ex. 617 at 171 (Furberg Dep. [discussing the extension portion of the study]). Consistent with the study authors, Dr. Bennett agreed that such data should be included in any statistical analysis. *See* Defs.' Ex. 793 at 93 (Bennett Dep.).

[42] Hr'g Tr. at 276-77; *see id.* at 105-06, 123; REF. MAN. at 380-81; Defs.' Ex. 612 at 23-24, 306-07 (Bennett Dep., *In re Bextra*); Defs.' Ex. 617 at 43, 46 (Furberg Dep.); Defs.' Ex. 627 at 53, 72 (Jewell Dep., *In re Bextra*); Defs.' Ex. 1899 at 626 (Furberg, AM. J. CARDIOL. 1984;53:626-27).

the studies. Pooled estimates relating to secondary hypotheses should only be considered if they have a clear pharmacological and biological foundation and even then should be interpreted conservatively." Defs.' Ex. 619 at 301 (Furberg & Morgan). Dr. Furberg even has written that "pooling of adverse effects is fraught with problems and is seldom feasible." *Id.* Indeed, it is telling that the FDA's 17-page, single-spaced decision memorandum regarding labeling for all NSAIDs did not mention any meta-analysis submitted to the FDA as a basis for its determinations. *See* Defs.' Ex. 664, *passim* (FDA Decision Mem.).

These reliability concerns are especially warranted with respect to Dr. Madigan's analyses because he has changed his methods repeatedly. In August 2008, Dr. Madigan first analyzed the Celebrex data for litigation. *See generally* Defs.' Ex. 707 (Madigan Rep., *Grutka v. Pfizer*). There, he collected and counted non-fatal events using a process that depended on the assistance of Dr. Baruch, who also allegedly analyzed fatal events by diagnosing causes of death based on one-line descriptions furnished to him by Dr. Madigan. *See* Defs.' Ex. 770 at 11, 17-21, 37-38, 71-72 (Baruch Dep.); Defs.' Ex. 622 at 44-45, 52-54 (Madigan Dep.).

In his March 2009 report in this litigation, Dr. Madigan continued to use Dr. Baruch's counts for non-fatal events. However, for fatal event, Plaintiffs' counsel directed Dr. Madigan to substitute the cause of death diagnoses of Dr. Baruch, the board-certified cardiologist, with new diagnoses by the non-cardiologist Dr. Furberg, who has not practiced medicine in thirty-five years and does not know the current diagnostic criteria for heart attack or pulmonary edema. *See* Defs.' Ex. 616 at 49 (Furberg Dep. *Haslam v. Pfizer*); Defs.' Ex. 617 at 151, 182-83 (Furberg Dep.); Defs.' Ex. 622 at 44-45, 65-67 (Madigan Dep.). Dr. Baruch's cause of death determinations were discarded even though Dr. Madigan had no methodological concerns with his work, *see* Hr'g Tr. at 569, which in itself renders Dr. Madigan's analyses unreliable.

Further, because Dr. Furberg disagreed with Dr. Madigan's endpoint definitions, he changed them without telling Dr. Madigan. *See id.* at 236-37; Section III.C *infra*.[43]

After Defendants identified numerous methodological problems with Dr. Madigan's analysis, he prepared a "supplemental" analysis less than two weeks before the hearing, which changed the event counts for a third time based partly on a Pfizer pooled analysis submitted to the FDA in 2005. *See* Hr'g Tr. at 571-72; Defs.' Ex. 1931 at 2-3 (Madigan Supp. Decl.). From his August 2008 report through his "supplemental" analysis, Dr. Madigan calculated in excess of 600 *p*-values. *See generally* Defs.' Ex. 707 (Madigan Rep., *Grutka v. Pfizer*); Defs.' Ex. 706 (Madigan Rep.); Defs.' Ex. 1931 (Madigan Supp. Decl.). Based on certain data collection errors, Dr. Madigan admitted that he would not submit the earlier version of his meta-analysis for peer review. *See* Hr'g Tr. at 529. Dr. Madigan also admitted it was not methodologically proper for Dr. Furberg to change some of Dr. Madigan's endpoint definitions, so he changed them back without the agreement of any other medical expert in the case – even though he admits that he is not qualified to define endpoints without the help of a clinician. *See id.* at 534, 542, 575-76.

None of Plaintiffs' medical experts reviewed Dr. Madigan's analysis; in fact, despite his extensive participation in the process, Dr. Furberg said that since he is not a biostatistician he would not put his reputation behind it. *See id.* at 177, 184, 369, 527-28. Dr. Madigan's submission of successive reports that each rely on materially different methodologies calls into

[43] Moreover, Dr. Madigan's August 2008 report analyzed the trials comparing Celebrex to other NSAIDs and concluded that Celebrex did *not* increase the risk of heart attacks and strokes compared to other NSAIDs. *See* Defs.' Ex. 707 at 17 (Madigan Rep., *Grutka v. Pfizer*). By contrast, in his March 2009 report, Dr. Madigan ignored the trials comparing Celebrex to other NSAIDs altogether because Plaintiffs' counsel instructed him to do so, *see* Hr'g Tr. at 460-61, even though NSAIDs were not believed to increase thrombotic risk during the relevant time. *See* Hr'g Tr. at 673-74. Dr. Madigan even had Dr. Furberg classify events from those trials *in this litigation* – but failed to incorporate them into his analysis. *See* Hr'g Tr. at 573-74, 674-75.

question the reliability of his opinions at the outset. *See In re Bausch & Lomb, Inc., Contact Lens Solution Prods. Liab. Litig.*, 2009 WL 2750462, at *12-14 (D.S.C. Aug. 26, 2009) (excluding Plaintiffs' expert in part because she filed several successive reports, which created a "moving target").

B. Dr. Madigan's Analyses Only Purport to Show Statistical Significance Before APC for Two of His Four Endpoint Hypotheses, and Only at the Highest Celebrex Doses, Based Primarily on a Handful of Events from a Single Trial.

In his March 2009 report and his "supplemental" analysis, Dr. Madigan employed "Hard CHD" as his primary endpoint, plus three secondary endpoints: myocardial thromboembolic, cardiovascular thromboembolic (which included strokes), and cardiovascular mortality. *See* Defs.' Ex. 706 at 4-5 (Madigan Rep.); *see generally* Defs.' Ex. 1931 (Madigan Supp. Decl.).[44] Dr. Madigan did *not* find that Celebrex was associated with a statistically significant increase in risk before APC for either the cardiovascular thromboembolic or cardiovascular mortality endpoints – at any dose, in any of his analyses – although he admits that his cardiovascular thromboembolic endpoint is a close approximation of the thrombotic conditions described in the boxed warning on the Celebrex label. *See* Defs.' Ex. 706 at 23-28, 30-35 (Madigan Rep.); Defs.' Ex. 1931 at Ex. 1 & 2 (Madigan Supp. Decl.); Hr'g Tr. at 549-50. Rather than basing his conclusions on his two endpoints that did *not* show statistical significance, Dr. Madigan instead sub-divided his Hard CHD and myocardial thromboembolic analyses by dose. Out of hundreds of comparisons in his report, it was only when he restricted his analysis to the highest doses in two of his four endpoints that he was able to create results that had a *p*-value of less than 0.05

[44] Dr. Madigan defines his Hard CHD endpoint as heart attacks and sudden cardiac death, while his myocardial thromboembolic endpoint includes heart attacks, angina pectoris aggravated, cardiac arrest, circulatory failure, myocardial ischemia, myocardial rupture post-infarction, ventricular tachycardia, coronary thrombosis and sudden death. *See* Defs.' Ex. 706 at 4-5 (Madigan Rep.). Neither endpoint includes stroke. *See id.*; *see also* Hr'g Tr. at 545.

before APC. *See* Defs.' Ex. 706 at 11-14, 18-21 (Madigan Rep.); Defs.' Ex. 1931 at Supp. Tables 5, 7, 11, 13, 35, 37, & 50 (Madigan Supp. Decl.); Hr'g Tr. at 509, 523, 862.

Further, by focusing only on 33 placebo-controlled trials completed before APC, Dr. Madigan confined his analysis to as few as 12 events, which predominantly occurred in the Alzheimer's 001 trial, with its well-recognized randomization failure. *See* Defs.' Ex. 1931 at Supp. Tables 6, 12-13 (Madigan Supp. Decl.); Defs.' Ex. 2022 at 4 & Table 1 (Packer Written Direct). Dr. Madigan's 2009 analyses also report that there were five Hard CHD events from that trial, which differs from his August 2008 report and directly conflicts with Dr. Furberg's interpretation of those events under oath. *Compare* Defs.' Ex. 707 at 8 (Madigan Rep., *Grutka v. Pfizer*) *with* Defs.' Ex. 1931 at Supp. Tables 2, 26 (Madigan Supp. Decl.); *see* Defs.' Ex. 617 at 407-09 (Furberg Dep.); Hr'g Tr. at 407 (Dr. Kronmal agreeing there were two thrombotic events in the trial). Had Dr. Madigan correctly counted those events, his results would not have been statistically significant. *See* Defs.' Ex. 622 at 175-76 (Madigan Dep.); Hr'g Tr. at 699-700.

In other words, only a portion of a portion of Dr. Madigan's analysis (only two of his four endpoints, and only at the highest Celebrex doses) shows nominally significant results. That portion is based on hundreds of comparisons and data from one flawed trial where Dr. Madigan miscounted the number of events and randomization did not evenly distribute patients on the basis of their cardiovascular risk factors before any patient took medications in the study.

C. The Endpoints on Which Dr. Madigan Relies Never Have Been Used in the Peer-Reviewed Literature to Evaluate the Thrombotic Safety of NSAIDs and Are Not Validated as Being Clinically Relevant by Any Medical Expert.

To make matters worse, the two endpoints on which Dr. Madigan relies do not measure heart attack risk, do not include stroke, never have been used by the peer-reviewed scientific and medical communities to analyze the thrombotic safety of NSAIDs in a clinical trial, differ from his endpoint methodology in the Vioxx litigation, and have not been validated as a clinically

valid thrombotic endpoint by any other medical expert. Indeed, the only remaining Plaintiffs' expert with any medical background at all, Dr. Furberg, had such problems with one of Dr. Madigan's endpoints that he unilaterally changed it and said he hoped never to see it again.

As a threshold matter, Dr. Madigan cannot explain the biological basis for his multiple endpoints or what they are supposed to measure, even though he agrees it is necessary for a hypothesis being tested to have some medical or biological underpinning. *See* Hr'g. Tr. at 533, 535, 551-52. He also admits that, as a statistician, he is not qualified to identify an appropriate endpoint and instead must rely on the advice of a qualified clinician (in this case, a cardiologist). *See id.* at 534, 542 (admitting he cannot identify any trial evaluating thrombotic risk where a cardiologist was not involved in endpoint selection). As a result, he does not claim that any of his endpoints validly measure heart attack risk, nor do they include stroke. *See id.* at 545, 551.

Dr. Madigan's inability to establish that either endpoint validly measures thrombotic risk is particularly striking because Dr. Madigan does not rely on any endpoint that was used by the FDA or in the peer-reviewed clinical trial literature to evaluate the thrombotic safety of NSAIDs during the relevant time. Dr. Madigan did not use a primary composite endpoint from any of the Pfizer-sponsored, peer-reviewed, and pre-specified analyses from clinical trials, including APC, PreSAP, or CABG-2; he did not use the thrombotic endpoint the FDA used in evaluating the CLASS and VIGOR studies; and he did not use the well-accepted APTC endpoint used in the Pfizer-sponsored White meta-analysis and by the FDA in 2005. *See* Hr'g Tr. at 564 (APC, PreSAP, or CABG-2); *id.* at 558 (no other NSAID literature using Hard CHD); *id.* at 559-60 (admitting he does not know if Hard CHD was used by Pfizer or anyone else to evaluate clotting effects in 1999); *id.* at 549-50 (admitting that his Hard CHD endpoint is not consistent with the boxed warning). In fact, there is no evidence in the record that, before APC, the FDA or any

other medical expert used either of Dr. Madigan's composite endpoints to analyze the thrombotic risk of an NSAID in a clinical trial or meta-analysis of clinical trials. *See* Hr'g Tr. at 558-60; Defs.' Ex. 622 at 230, 245, 255-57 (Madigan Dep.). Further undermining the reliability of his analysis, Dr. Madigan discarded the endpoint he used as a paid expert in the Vioxx litigation, which included stroke and unstable angina, even though he could not identify any published meta-analyses where the authors used different endpoints to analyze the same safety issue for two medications in the same class. *See* Hr'g Tr. at 557-58, 877-78; Defs.' Ex. 622 at 238-239 (Madigan Dep.).

Perhaps not surprisingly, none of Plaintiffs' medical experts validated Dr. Madigan's endpoint selections. Dr. Madigan claimed that he spoke with Dr. Baruch about his endpoints, but Dr. Baruch had no such recollection, and Dr. Madigan did not mention Dr. Baruch in any of his reports. *See* Defs.' Ex. 770 at 41-43 (Baruch Dep.). The board-certified cardiologist Dr. Baruch also was not familiar with the "Hard CHD" endpoint, could not define it, and could not recall ever using that endpoint himself or even reading a published paper that used it. *See id.* at 43-44, 46-48. By contrast, he testified that he predominantly looks at heart attacks, strokes, and cardiovascular deaths – that is, the APTC endpoint – in his analysis of Bextra and Celebrex. *See id.* at 74-75, 147. Dr. Madigan also never spoke with the blood clotting and thrombosis expert Dr. Bennett, Dr. Furberg, or Dr. Kronmal about his endpoints. *See* Hr'g Tr. at 557, 565, 878-79, 881; Defs.' Ex. 617 at 428-29 (Furberg Dep. [testifying that the Hard CHD endpoint is too narrow]). Thus, despite Dr. Madigan's acknowledgment that he lacks the expertise to define a clinically valid endpoint, no other Plaintiffs' expert has validated his endpoint choices.

Dr. Madigan's "Hard CHD" endpoint also suffers from other methodological problems. Drs. Kronmal and Madigan admitted that researchers use different variations of that endpoint.

See Hr'g Tr. at 402, 561. In fact, while Dr. Madigan claims that Hard CHD was used in a trial of a cholesterol-lowering medication, that trial used a different definition than the one Dr. Madigan outlined in his report.[45] In light of these inconsistencies, it is inexplicable that Dr. Madigan did not provide Drs. Baruch or Furberg with a written definition of the composite endpoint – he simply assumed that they would employ an identical and valid version of that endpoint, but he does not know what definition they applied. *See* Hr'g Tr. at 561-63. Dr. Madigan also admitted that Pfizer never used his Hard CHD endpoint in its own analyses. *See id.* at 546, 551.

With respect to his myocardial thromboembolic endpoint, Dr. Madigan claims that he derived that endpoint from the 2005 Pfizer pooled analysis because he was trying to replicate what Pfizer knew. Yet Dr. Madigan admits that the myocardial thromboembolic endpoint was only a *component* of Pfizer's primary endpoint in its analysis and that Pfizer never claimed that component was clinically meaningful or a reliable measure of thrombotic risk. *See* Hr'g Tr. at 872-73. In contrast to the portion relied on by Dr. Madigan, Pfizer's *complete* primary endpoint is consistent both with the boxed label and Dr. Furberg's thrombotic definition and did not show a statistically significant increase in thrombotic risk. *See id.* at 549-50, 872-73. In other words, unlike in his Vioxx litigation work, here Dr. Madigan departed from Pfizer's primary endpoint and instead based his opinion on a "myocardial thromboembolic" portion of Pfizer's primary endpoint, even though Dr. Madigan was unable to cite any precedent establishing the validity of that portion in the peer-reviewed medical literature. *See* Hr'g Tr. at 560, 872-73.

[45] *See* Hr'g Tr. at 558-59, 561-62; *compare* Defs.' Ex. 710 at 1149 (Sever et al., LANCET 2003;361:1149-58 [heart attacks and fatal coronary heart disease ("CHD")]) *with* Defs.' Ex. 706 at 4 (Madigan Rep. [heart attacks and *sudden* cardiac death ("SCD")]); *see also* Defs.' Ex. 15 at 4 (ASCOT Endpoint Manual [defining Hard CHD as all CHD deaths, including non-SCDs]); Hr'g Tr. at 560 (admitting he does not know the difference between SCD and fatal CHD).

Dr. Madigan's myocardial thromboembolic endpoint also "troubled" Dr. Furberg, who was "taken aback" because he never had seen it in his 40 years of clinical research experience. Hr'g Tr. at 236-38; *see* Defs.' Ex. 617 at 396 (Furberg Dep.). In fact, he testified that he never wants to see it again "because it doesn't make sense to me as a physician, as a scientist." Defs.' Ex. 617 at 429-30 (Furberg Dep.). Dr. Furberg even changed Dr. Madigan's endpoint definition in the middle of his diagnosis process – without informing Dr. Madigan. *See id.* at 397-98; Hr'g Tr. at 237. At the hearing, Dr. Madigan admitted that changing an endpoint definition in the midst of an analysis is not methodologically valid and that he "would probably not" submit a meta-analysis for peer review where a reviewer changed the definition midstream. Hr'g Tr. at 575-76 (noting further that he was "uncomfortable" with Dr. Furberg changing the definition and that such a change is "undesirable"); *see id.* at 575 ("I'd rather that didn't happen."). Rather than seeking a second opinion from a cardiologist or thrombosis expert, Dr. Madigan changed the definition back without the advice of any medical expert at all.

With no biological basis for his various endpoint hypotheses; no precedent for the endpoints in the peer-reviewed NSAID literature, in an FDA analysis, or in his own prior litigation work; no cardiologist or other medical expert to validate his endpoint selections; and nothing indicating that Pfizer used the "myocardial thromboembolic" portion of its primary endpoint as a reliable measure of thrombotic risk, there is nothing to suggest that Dr. Madigan's multiple alternative endpoint methodologies have a reliable clinical basis. Dr. Madigan's use of so many endpoints – and the hundreds of calculations he performs across those endpoints – calls into question the reliability of his analysis. *See* Hr'g Tr. at 538. In contrast, using the thrombotic endpoint definitions of the FDA, Plaintiffs' own medical experts, NSAID clinical trials before 2004, Dr. Madigan's Vioxx endpoint, or the only published meta-analysis recognized as

authoritative by any of Plaintiffs' experts (Kearney), there is no reliable basis to conclude that an increased risk of thrombotic events was established prior to December 16, 2004.

D. Dr. Madigan's Analyses Use Unreliable Data Collection and Classification Procedures That Never Have Been Used in the Peer-Reviewed Literature.

In addition to using endpoints not used in NSAID clinical trials or literature before APC, Dr. Madigan also employed an unreliable and undocumented data collection process that never has been used outside of this litigation. He searched for events using only trials for which SAS data files were available, even though he cannot identify any published meta-analysis using that approach other than his own upcoming paper based on his Vioxx litigation work. *See* Hr'g Tr. at 567 ("I have no idea whether this happens, has happened routinely in the past or not.").

While Dr. Madigan claims he relied on Dr. Baruch to help him identify all relevant non-fatal events, Dr. Baruch had no recollection of that and was not subject to cross-examination at the hearing. *See* Defs.' Ex. 770 at 53-59 (Baruch Dep.). Dr. Baruch also was not provided with the available medical information for each non-fatal event, an omission Dr. Furberg criticizes outside of litigation. *See* Defs.' Ex. 617 at 101-103 (Furberg Dep. [criticizing ADAPT results as "soft" because events were not reviewed by cardiovascular experts]); note 23 *supra*. Moreover, because neither Dr. Madigan nor Dr. Baruch wrote down the patient identification numbers that supposedly correspond with each non-fatal event, there is no way to replicate or validate their methodology, a prerequisite for reliability under *Daubert*.[46] *See Innis Arden Golf Club v. Pitney*

[46] Contrary to Dr. Madigan's claims at the hearing, Dr. Baruch also: (1) does not recall how he obtained the dictionary he used to identify relevant non-fatal events or what terms he used, *see* Defs.' Ex. 770 at 57-58 (Baruch Dep.); (2) could not remember whether he discussed the terms with Dr. Madigan, *see id.* at 41-43; and (3) could not validate the terms listed on the document that Plaintiffs represent as reflecting the search terms that Dr. Baruch endorsed, which led Dr. Baruch to admit he could offer no proof or validation that any of the terms Dr. Madigan used to search for "Hard CHD" or his other endpoints were valid ones to use. *See id.* at 163-65.

Bowes, Inc., 629 F. Supp. 2d 175, 190 (D. Conn. 2009) (excluding expert opinion because his samples and data packages no longer existed and thus "[d]efendants could not attempt to validate [his] methods even if he could specifically say what he considered"). Significantly, Dr. Madigan's unprecedented, undocumented process also failed to extract a number of relevant fatal events. *See* Hr'g Tr. at 678-79; Defs.' Ex. 770 at 134-36 (Baruch Dep.); *see also* Defs.' Reply Br., App. Fig. 16.

For the fatal events Dr. Madigan did extract, prior to his "supplemental" analysis, he provided Dr. Baruch (and later Dr. Furberg) only the one-line descriptions from the SAS data files, rather than the detailed patient narratives necessary to interpret causes of death. *See* Hr'g Tr. at 566; *id.* at 679-82 (providing examples where SAS data file snippets were far less reliable than full narratives); Defs.' Ex. 2023 at 18-19 (Dr. Packer Demonstratives). Dr. Furberg admitted that the process was "challenging" and said that trying to divine a cause of death with so little information was a "weakness" of the process. Hr'g Tr. at 186-89. He also admitted that he never had taken part in such a rushed process based on so little information in his entire career. *See* Defs.' Ex. 617 at 344, 385-86 (Furberg Dep.). In fact, Dr. Furberg *assumed* Dr. Baruch made his cause of death determinations based on all available medical information – not one-line phrases from the SAS data files. *See* Hr'g Tr. at 188. Dr. Madigan also failed to specify – or ask Drs. Baruch or Furberg to provide – uniform diagnostic criteria for any of the events included in his composite endpoints or to establish a procedure to resolve any differences of opinion, even though Dr. Furberg could not identify the current criteria for diagnosing a heart attack. *See* Hr'g Tr. at 191; Defs.' Ex. 617 at 182-183 (Furberg Dep.).

Dr. Madigan's failure to use patient narratives, the lack of pre-specified diagnostic criteria, and the lack of a procedure for resolving differences between Drs. Baruch and Furberg

stand in stark contrast to peer-reviewed, published methods that are used to ensure that event counts which form the basis of a statistical analysis are reliable. *See* Defs.' Ex. 617 at 417-18 (Furberg Dep.); Hr'g Tr. at 192; Section I.B *supra*. In light of this unreliable and unprecedented process, there were numerous differences in how the events were categorized for Dr. Madigan's analysis, including between Drs. Baruch and Furberg, between Drs. Baruch and Furberg and the published literature, and even within Dr. Furberg's analysis.[47]

For the first time, in the context of Dr. Madigan's "supplemental" analysis, he provided Drs. Baruch and Furberg with detailed patient narratives, but only for four patients out of the approximately 1,800 available patient narratives in the Celebrex trials. *See* Hr'g Tr. at 566-67. Dr. Madigan could offer no coherent explanation for why he gave Drs. Baruch and Furberg only four patient narratives when more were available for their review. *See id.* at 579.

Dr. Madigan's "supplemental" analysis also relied – for the first time – on event counts from the 2005 Pfizer pooled analysis, even though he would not confirm that Pfizer's analysis was reliable. *See* Hr'g Tr. at 653. Dr. Madigan agrees the interpretation of cardiovascular data should be performed by medical doctors with expertise in cardiology, but he admits that the 2005 Pfizer pooled analysis was performed by Pfizer employees who were not cardiologists. *See id.* at 870, 873-74. Dr. Madigan also acknowledges that researchers attempting to interpret whether a patient experienced a specified condition should be blinded to the treatment the subjects received (in fact, he cannot name any high quality meta-analysis in which reviewers were not blinded),

[47] *Compare* Defs.' Ex. 706 at 30 (Madigan Rep. [14 cardiovascular deaths for APC]) *with* Defs.' Ex. 743 at 1030-31 (Solomon S., et al., CIRCULATION 2006;114:1028-1035 [12 cardiovascular deaths]); *see* Defs.' Ex. 793 at 77-78 (Bennett Dep. [stating it would not be appropriate to change event counts from the cardiovascular safety committees in APC and PreSAP]); Hr'g Tr. at 193 (Dr. Furberg has no reason "whatsoever" to doubt the published APC counts); Defs.' Ex. 617 at 407-09 (Furberg Dep. [only one Hard CHD death in Alzheimer's 001]); Hr'g Tr. at 684-86.

but he concedes that the Pfizer scientists knew which medication the subjects were taking. *See id.* at 873-75. As a result, Pfizer pointed out the limitations inherent in the analysis at the time. *See* Defs.' Ex. 1057 at 25, 27, 32-33 (Celecoxib Meta-Analysis Rep., Apr. 25, 2005 [noting further that the results were driven by the Alzheimer's 001 trial, which suffered from randomization failure, and that Pfizer made no multiple comparison adjustments]). Despite the limitations of that analysis, its primary endpoint – which is consistent with Dr. Furberg's definition of a thrombotic event and the boxed label – still did not show a statistically significant increase in thrombotic events. *See id.* at 32; Hr'g Tr. at 869, 872. Moreover, Dr. Madigan missed events that were included in that analysis.[48] Dr. Madigan's reliance on the unblinded, non-published, non-peer reviewed pooled analysis by non-cardiologists is perplexing given that Dr. Madigan ignored the Pfizer-sponsored, peer-reviewed, cardiologist-adjudicated versions of the same data. *See* Hr'g Tr. at 683-84; note 8 *supra* (citing the White meta-analyses).

Moreover, as Plaintiffs' experts acknowledge, pooled analyses with multiple comparisons are unreliable since it is impossible to know the true error rate for such comparisons. *See* Section I.C.3 *supra*. Yet Dr. Madigan made no adjustment to his statistical significance thresholds, even though only a handful of the hundreds of *p*-values he calculated fell below thresholds typically used for single comparisons, and even then for only two of his endpoints. *See* Hr'g Tr. at 537. For all these reasons, Dr. Madigan's meta-analysis is unreliable and inadmissible under *Daubert*.

[48] *See* Hr'g Tr. at 675-77, 735-36; *compare* Defs.' Ex. 1399 at 70 (COXA-0508-249 Study Rep. [patient 0005-0048]); Defs.' Ex. 1376 at 118 (635-IFL-0508-003 Study Rep. [patient 2009-2042]) *with* Defs.' Ex. 320 at 5 (Table 2); Defs.' Ex. 1057 at 32 (Table 2), 40 (Table 8), 84 (Table 5.1.2.1); Defs.' Ex. 1931, Supp. Table 10; Defs.' Ex. 2013 at 16 (Table 2), 28 (Table 10).

IV. THE CABG SURGERY TRIALS DO NOT ESTABLISH THAT ORAL BEXTRA INCREASES THE RISK OF THROMBOTIC EVENTS.

The totality of evidence today establishes that arthritis patients taking common doses of Bextra are at no greater risk of heart attacks and strokes than those taking other NSAIDs like Motrin or Aleve or no NSAID at all. *See* Defs.' Ex. 632 at 83 (Ad. Comm. Tr., Feb. 17, 2005 ["With [Bextra] . . . the information we have at this time suggests that the risk is not increased at doses of 20 mg or less."]). Only the CABG surgery trials, experimental studies involving high doses of intravenous parecoxib – a medication never approved in the United States for use outside of clinical trials – followed by high doses of oral Bextra pills in patients immediately after undergoing CABG surgery, showed a statistically significant increase in the risk of thrombotic events. *See* Hr'g Tr. at 179-80, 425; Defs.' Ex. 686 at 28-29 (Bextra Integrated Summary of Safety). Even the CABG surgery trials did not show a statistically significant increase in risk for oral Bextra pills taken alone in the absence of intravenous parecoxib. For the reasons outlined below, Plaintiffs' experts' opinions about the thrombotic safety of oral Bextra based on the high-dose CABG surgery trials involving parecoxib are unreliable.

A. Plaintiffs' Experts Ignore the Differences Between Intravenous Parecoxib and Oral Bextra Pills.

Plaintiffs' experts cite the CABG studies as evidence of an increased thrombotic risk for patients taking Bextra, even in the absence of parecoxib. Yet Plaintiffs' experts admit that they *ignored* the different cardiovascular effects seen with intravenous parecoxib. For example, Dr. Furberg claimed in his report that Bextra increases blood pressure, which he said raised an important cardiovascular safety issue. *See* Defs.' Ex. 610 at 32-33, 42-43 (Furberg Rep.). At the same time, he offers no explanation for the fact that parecoxib significantly *decreased* blood pressure in both CABG trials – the exact opposite effect on blood pressure that he associates with the use of Bextra pills. *See* Defs.' Ex. 632 at 507 (Ad. Comm. Tr., Feb. 16, 2005); Defs.' Ex.

690 at 3 (FDA Ltr.). Dr. Furberg's failure to consider such differences is striking since he has opined – outside of litigation and as a retained expert for plaintiffs – that different forms of administration of a medication can have different effects.[49] Where an expert fails to consider whether his assumptions are correct – here, the assumption that Bextra and parecoxib have the same effect – opinions based on those assumptions must be excluded. *See In re Baycol Prod. Litig.*, 532 F. Supp. 2d 1029, 1046 (D. Minn. 2007); *Hamilton v. Emerson Elec. Co.*, 133 F. Supp. 2d 360, 371-72 (M.D. Penn. 2001).

B. Plaintiffs' Experts Lack the Expertise to Evaluate the Unique Physiology of CABG Surgery Patients.

Plaintiffs' remaining experts are not qualified to evaluate the unique physiology of CABG surgery patients and therefore lack the necessary foundation to opine that the CABG trials provide a reliable basis for thrombotic risk opinions about oral Bextra pills. Plaintiffs did not produce Dr. Bennett at the hearing, even though he is their blood clotting and thrombosis expert, their only expert who professes to understand how these medications work in the body, and has studied the function of blood platelets in the context of CABG surgery. Dr. Bennett readily admits that the physiology of CABG patients is so unique that it is unreliable to apply the results of the CABG trials to arthritis patients. "[T]his is a completely different setting. So I'm not sure you could legitimately or logically extrapolate from going on a bypass machine to treating swollen joints." Defs.' Ex. 793 at 137 (Bennett Dep.); *see id.* at 120-23. As a result, Dr. Bennett concedes that it is improper to "extrapolate the findings from the CABG data to

[49] *See* Hr'g Tr. at 265-68 (discussing examples where injectable forms of medications exerted different clinical effects than oral forms); Defs.' Ex. 197 at 181-82 (FUNDAMENTALS); Defs.' Ex. 623 at IV-18 (Furberg, CLIN. CARDIOL. 2000;23;7 Suppl. 4:IV15-19); Defs.' Ex. 616 at 362-65 (Furberg Dep., *Haslam v. Pfizer*). Plaintiffs' other experts also failed to consider differences between the effects of parecoxib and oral Bextra pills. *See* Defs.' Ex. 621 at 6, 317-18 (Kronmal Dep.); Defs.' Ex. 793 at 123-24, 219 (Bennett Dep.).

people taking the drug in the real world for arthritis" – thereby rejecting the very foundation of Plaintiffs' other experts' opinions. Defs.' Ex. 612 at 328-29 (Bennett Dep., *In re Bextra*); *see* Defs.' Ex. 793 at 124 (Bennett Dep.).

As for Plaintiffs' witnesses who did testify at the hearing regarding Bextra, Dr. Furberg admitted that he did not even consider the physiological differences between CABG surgery patients and arthritis patients because he is not qualified to do so.[50] Again, Dr. Furberg's failure to consider the impact of CABG surgery on patients' physiology in this litigation stands in stark contrast to his statements outside the courtroom, where he has admonished researchers repeatedly that they must take great care before applying clinical trial results beyond the population from which the trial was drawn. *See* Section I.C.4 *supra*. In fact, Dr. Furberg has admitted that surgical populations differ from outpatients. *See* Defs.' Ex. 632 at 527-28 (Ad. Comm. Tr., Feb. 16, 2005). Notably, Dr. Furberg admitted as much at the hearing when he distinguished the hypotension safety data in the CABG trials on the grounds that CABG surgery is a "*totally different setting*" and a "very complex surgical procedure" in which it is "very hard to tease up what is the drug effect, what is the effect of the procedure and all the other drugs people are getting in conjunction with surgery." Hr'g Tr. at 258-60 (emphasis added). As for Dr. Kronmal, Plaintiffs concede that he is not qualified to compare the physiology of patients undergoing CABG surgery to that of patients in the general arthritis population, claiming that he

[50] *See* Hr'g Tr. at 263-64; Defs.' Ex. 617 at 246-47 (Furberg Dep.); Defs.' Ex. 616 at 192 (Furberg Dep., *Haslam v. Pfizer*). Dr. Furberg is not an expert on cardiovascular anesthesiology or on how CABG surgery affects the body's clotting and circulatory systems, and he never has performed the surgery and cannot describe how it is performed. *See* Hr'g Tr. at 264-65; Defs.' Ex. 616 at 60-61 (Furberg Dep., *Haslam v. Pfizer*). Defendants' experts Drs. Frank Sellke and Barry Massie did analyze the differences between parecoxib and oral Bextra pills and CABG surgery and arthritis patients, which Plaintiffs did not rebut. *See* Hr'g Tr. at 589-94; Defs.' Ex. 2020 at 13, 22-25 (Sellke Written Direct); Defs.' Ex. 2026 at 37-38 (Massie Written Direct).

only evaluated Bextra from a "statistical perspective." Pls.' Opp. at 90; *see* Defs.' Ex. 621 at 124-27 (Kronmal Dep.); Defs.' Ex. 627 at 304-305 (Jewell Dep., *In re Bextra* [admitting one needs more than statistical expertise to generalize results]).

C. Dr. Furberg's Attempts to Combine the Parecoxib CABG Surgery Trials with the Bextra Arthritis Trials Have Been Rejected by His Peers.

Based on the dramatic differences ignored by Plaintiffs' experts, the medical and scientific communities have rejected attempts to combine data from trials involving intravenous parecoxib with arthritis trials involving oral Bextra. When Dr. Furberg tried to publish a meta-analysis combining Bextra arthritis data with the parecoxib CABG trials, all journals to which he submitted the article rejected his methodology. *See* Hr'g Tr. at 272-73; Defs.' Ex. 616 at 313-14, 320-21, 351-52 (Furberg Dep., *Haslam v. Pfizer*); Defs.' Ex. 809 at 1 (Transaction History). Dr. Furberg could not identify any meta-analyses that found the data similar enough to combine in a single analysis. *See* Hr'g Tr. at 273-74. In a published editorial, which treated the CABG data as unique, Dr. Furberg even questioned whether the results from the CABG surgery setting were applicable to an arthritis population taking lower doses. *See* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249 ["It is currently unclear to what degree such risk extends to patients treated chronically with lower doses for arthritis. . . ."]).

Plaintiffs cite comments by the FDA in deciding – as a precautionary public health matter – that it is reasonable to *assume* that Bextra "does not differ from the other COX-2 selective agents with regard to increased [cardiovascular] risk with chronic use pending the availability of data from long-term controlled clinical trials that would indicate otherwise." Defs.' Ex. 664 at 9 (FDA Decision Mem.). The FDA's public health assumption is insufficient to save Plaintiffs' experts' opinions, however, for three reasons. First, the FDA repeatedly questioned the applicability of the CABG surgery trials to arthritis patients taking oral Bextra. In 2001, when

Pfizer asked the FDA to approve language describing the CABG data in the Bextra label, Defs.' Ex. 694 at 5 (Ltr. from Searle to FDA), the FDA rejected the proposed language and removed all mention of the CABG data in the Bextra label because the results were not reliable to evaluate the safety of Bextra pills when used as approved.[51] In its 2005 memorandum, the FDA stated that it was "difficult to know how to extrapolate the findings from the parecoxib/[Bextra] CABG trials to the chronic use situation given the significant physiologic and traumatic impact on the coronary vasculature during and following CABG surgery, and the systematic pro-inflammatory response resulting from heart-lung bypass." Defs.' Ex. 664 at 9 (FDA Decision Mem.). Numerous experts at the Advisory Committee hearing made similar comments.[52] Moreover, the FDA approved Bextra but not parecoxib, recognizing that the medications have different effects, particularly with respect to potentially life-threatening hypotension – a side effect Plaintiffs' experts ignored in reaching their opinions. *See* Defs.' Ex. 690 at 2-3 (FDA Ltr.).

Second, Plaintiffs ignore case law establishing that precautionary public health assumptions made by regulatory agencies outside of litigation are based on different standards and thus are insufficient to support an expert opinion subject to *Daubert*. *See* note 28 *supra*.

Third, whether an opinion is supported by a reliable foundation is precisely the question *Daubert* is meant to address, and expert opinions must be excluded when the methodology is unreliable or unfounded. *See McClain v. Metabolife Int'l, Inc.*, 401 F.3d. 1233, 1244-45 (11th

[51] *See* Defs.' Ex. 695 at 2-3 (Minutes of Oct. 24, 2001 NDA Meeting); Defs.' Ex. 693 at 3 (Medical Officer Review, Nov. 7, 2001); Defs.' Ex. 824 at 60-62, 80 (Goldkind Dep., *In re Bextra*); Defs.' Ex. 687 at 4-19 (Bextra Initial Approval Letter and Label, Nov. 16, 2001).

[52] *See* Defs.' Ex. 666 at 12 (Ad. Comm. Minutes ["[T]he Committee felt that the evidence was very limited and it is difficult to extrapolate [the CABG results] to a real life setting."]); Defs.' Ex. 632 at 300 (Ad. Comm. Tr., Feb. 18, 2005 ["[I]n the case of cardiopulmonary bypass, I really do think that is a very different kettle of fish."]); *id.* at 530 (Ad. Comm. Tr., Feb. 16, 2005 ["The CABG population is very different, very much a pro-inflammatory population."]).

Cir. 2005); *Hollander v. Sandoz Pharms. Corp.*, 289 F.3d 1193, 1208 (10th Cir. 2002). Because the "analytical gap" between the CABG surgery trial data and Plaintiffs' experts' opinions is so great, their CABG opinions do not survive review under *Daubert*. *In re Bextra I*, 524 F. Supp. 2d at 1171, 1180-81 (listing "whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion" as a factor in assessing reliability and excluding testimony where the gap between the data and the expert's conclusion was "simply too great to make the opinion admissible").

V. PLAINTIFFS' MOTION TO EXCLUDE DR. WEI'S META-ANALYSIS

Plaintiffs bear the burden of proof that there was reliable, statistically significant evidence of an increased risk of thrombotic events with Celebrex and Bextra prior to December 16, 2004. Defendants proffered Dr. Wei's meta-analysis only in rebuttal to Plaintiffs' experts' opinions and do not rely on it as a basis for their motion to exclude those opinions. Because the Court grants Defendants' motion, there is no need to rule on the admissibility of Dr. Wei's meta-analysis, and Plaintiffs' motion to exclude it is DENIED as moot.

In the alternative, Plaintiffs' motion is DENIED on the merits. Dr. Jewell's criticisms echo the ones he made in the Celebrex product liability litigation, which both courts summarily rejected. *See In re Bextra I*, 524 F. Supp. 2d at 1184; *In re Bextra II*, 2008 N.Y. Misc. LEXIS 720 at *51. Moreover, unlike Defendants' motion, Plaintiffs' motion is based not on undisputed principles or Dr. Wei's concessions regarding proper methods, but on the criticisms of their own expert Dr. Jewell, who never has performed a meta-analysis outside of litigation. *See* Defs.' Ex. 627 at 50 (Jewell Dep., *In re Bextra*); Defs.' Ex. 628 at 13-23, 25-27 (Jewell Dep.). Plaintiffs thus have resorted to a "battle of the experts" that would require the Court to weigh the respective experts' opinions, which *Daubert* does not permit. *Columbus Drywall & Insulation, Inc. v. Masco Corp.*, 2009 WL 856306, at *6 (N.D. Ga. Feb. 9, 2009).

Finally, Dr. Jewell's criticisms lack merit for the following reasons: (1) Plaintiffs do not contend that Dr. Wei is not qualified to conduct a meta-analysis, *see* Defs.' Wei Opp'n at 14-15; (2) Dr. Wei's meta-analysis employed the same level of intellectual rigor as meta-analyses conducted outside of litigation, *see id.* at 15-18; (3) Dr. Wei used several alternative statistical methods (known as performing "sensitivity analyses"), which showed that using different but equally accepted methodological approaches – including ones that account for the different lengths of the trials he included – would not affect his results, *see id.* at 18-25; Defs.' Ex. 742 at 23 (Wei Rep.); Defs.' Ex. 530 at 18 (Wei Decl.); Hr'g Tr. at 753; (4) Plaintiffs' experts' criticisms of Dr. Wei's methods go to the weight of his opinions, rather than their admissibility, *see* Defs.' Wei Opp'n at 26-33, 43-44; (5) using the numbers of events divided by the number of patients as opposed to patient-years is well-recognized in the NSAID literature, other literature, and even Dr. Furberg's own Bextra meta-analysis outside of litigation;[53] and (6) Dr. Jewell's criticism is inconsistent with Dr. Bennett's hypothesis of how Bextra and Celebrex work in the body, according to which any increased thrombotic risk should be present immediately and remain constant over time. *See* Defs.' Ex. 612 at 98-101 (Bennett Dep. *In re Bextra*).

CONCLUSION

Based on the foregoing findings and conclusions, the Court hereby GRANTS the Defendants' motion in its entirety and DENIES the Plaintiffs' motion.

SO ORDERED.

[53] *See* Defs.' Ex. 629 at 249 (Furberg et al., CIRCULATION 2005;111:249); Defs.' Ex. 616 at 296 (Furberg Dep., *In re Haslam*); Defs.' Ex. 530 at 6 (Wei Decl. [citing articles]); Defs.' Ex. 703 at 763 (Chen et al., PHARMACOEPIDEMIOL. DRUG SAF.; 16:762-72); Defs.' Ex. 792 at 133 (Caldwell et al., J. ROYAL SOC. MED. 2006;99:132-40).

Respectfully Submitted,

New York, New York
December 4, 2009

CADWALADER, WICKERSHAM & TAFT LLP

By: /s/ Gregory A. Markel

Dennis J. Block
Gregory A. Markel
Jason M. Halper
One World Financial Center
New York, New York 10281
Telephone: 212-504-6000
Facsimile: 212-504-6666

Of Counsel:

Loren H. Brown
Christopher G. Campbell
Matthew A. Holian
DLA PIPER LLP (US)
1251 Avenue of the Americas
New York, NY 10020-1104
Telephone: 212-335-4846
Facsimile: 212-335-4501

Charles Q. Socha
TUCKER ELLIS & WEST LLP
Metropoint 1, Suite 1325
4600 S. Ulster Street
Denver, CO 80237
Telephone: 720-897-4404
Facsimile: 720-222-5242